JOURNAL OF APPLIED PHYSICS

JOURNAL OF APPLIED PHYSICS

Vol. 25, No. 2 FEBRUARY, 1954

In This Issue

	Page
Transient-Flow Method for Determination of Thermal Constants of Insulating Materials in Bulk. Part I. Theory — J. H. Blackwell	137
Relation between Stress-Relaxation Studies and Dynamic Properties of Polyisobutylene — E. Catsiff and A. V. Tobolsky	145
Ferrites for Microwave Circuits and Digital Computers — E. Albers-Schoenberg	152
Synthesis of X-Ray Spectrometer Line Profiles with Application to Crystallite Size Measurements — Leroy Alexander	155
Rocket Measurements of Upper Atmosphere Ambient Temperature and Pressure in the 30- to 75-Kilometer Region — H. S. Sicinski, N. W. Spencer, and W. G. Dow	161
Penetration of an Electromagnetic Wave into a Ferromagnetic Material — A. Papoulis	169
Ferromagnetic Spinels with Rectangular Hysteresis Loops — I. J. Hegyi	176
Coupling of Modes of Propagation — J. R. Pierce	179
Electron Diffraction Study of the Structures of Antimony Electrodeposited on Antimony Cleavage Face — Ling Yang	184
Turbulent Processes as Observed in Boundary Layer and Pipe — G. B. Schubauer	188
Analysis and Purification of Rare Gases by Means of Electric Discharges — R. Riesz and G. H. Dieke	196
Ultra-High Vacuum. II. Limiting Factors on the Attainment of Very Low Pressures — D. Alpert and R. S. Buritz	202
Fast Time Analysis of Intermittent Point-to-Plane Corona in Air. I. The Positive Point Burst Pulse Corona — M. R. Amin	210
Parallel-Connected Magnetic Amplifiers — S. H. Chow	216
Current Noise in Thermistor Bolometer Flakes — James J. Brophy	222
Diffusion of Pulsed Currents in Conductors — L. M. Vallese	225
Effect of Organic Vapor on the Secondary Emission of Phosphors — P. H. Dowling and J. R. Sewell	228
Design and Calibration of Stagnation Temperature Probes for Use at High Supersonic Speeds and Elevated Temperatures — Eva M. Winkler	231
Analysis of Quantitative Pole-Figure Data — C. G. Dunn	233
Plastic Deformation of Ammonium Dihydrogen Phosphate — Paul L. Smith and Edward I. Salkovitz	237
X-Ray Noise Observation Using a Photoconductive Pickup Tube — A. D. Cope and A. Rose	240
On the Focusing of High-Current Electron Beams — George R. Brewer	243
Solution of Nonlinear Differential Equations by the Reversion Method through the First Thirteen Terms — George I. Cohn, and Bernard Saltzberg	252
Distortion of a Crystal by Point Imperfections — J. D. Eshelby	255
Approximate Method of Evaluating Integral Transforms — Armen H. Zemanian	262
Letters to the Editor:	
New Magnetostrictive Material — H. E. Stauss and G. Sandoz	267
Change of the Isotopic Abundance Ratio within a Sphere Due to Diffusion — J. T. Bracken and F. E. Senftle	267
Radioactive-Shadowing Technique and Note on the Deposition of Metal Crystals from Atomic Beams — M. Simnad, Ling Yang, and G. M. Pound	268
Comments on Diffraction of Electromagnetic Waves by an Aperture in a Large Screen — J. H. Crysdale	269
Momentum Transfer in Sputtering by Ion Bombardment — Gottfried K. Wehner	270
Rotation of the Microwave Polarization in Liquid Crystals — Glenn A. Mann and R. D. Spence	271
Cover Photograph	272
Index to Advertisers	xxiv

ELMER HUTCHISSON, *Editor*

EARLE C. GREGG, *Assistant Editor*

Board of Editors
MAURICE A. BIOT
RAY E. BOLZ
WALLACE J. ECKERT
ROBERT D. HEIDENREICH
LOUIS A. PIPES
ROMAN SMOLUCHOWSKI
W. R. SMYTHE
ARTHUR V. TOBOLSKY
ALVIN M. WEINBERG

Publication of the American Institute of Physics

OFFICERS
GEORGE B. HARRISON, *Chairman*
GEORGE B. PEGRAM, *Treasurer*

GOVERNING BOARD
J. W. BUCHTA
KARL K. DARROW
G. J. DIENES
S. A. GOUDSMIT
GEORGE B. HARRISON
FREDERICK V. HUNT
DEANE B. JUDD
HUGH S. KNOWLES
C. C. LAURITSEN
WILLIAM F. MEGGERS
PHILIP M. MORSE
BRIAN O'BRIEN
GEORGE B. PEGRAM
I. I. RABI
DUANE ROLLER
FREDERICK SEITZ
WILLIAM SHOCKLEY
RICHARD M. SUTTON
J. H. VAN VLECK
MARK W. ZEMANSKY

Consultant on Printing
MELVIN LOOS

ADMINISTRATIVE STAFF
HENRY A. BARTON, *Director*
WALLACE WATERFALL, *Executive Secretary*
THEODORE VORBURGER, *Advertising Manager*
RUTH F. BRYANS, *Publication Manager*
EDITH I. NEFTEL *Circulation Manager*
KATHRYN SETZE *Assistant Treasurer*

The JOURNAL OF APPLIED PHYSICS, published monthly at Prince and Lemon Sts., Lancaster, Pa., is devoted to physics in its role as the science basic to other natural sciences and to the arts and industries. Previous to 1937 this publication was known as PHYSICS. The JOURNAL OF APPLIED PHYSICS publishes editorials and reviews, as well as technical papers of applied physics. Articles appearing in it are indexed regularly, according to their field, by Physics Abstracts, Chemical Abstracts, Engineering Index, Inc., Industrial Arts Index, etc.

Manuscripts should be submitted to Elmer Hutchisson, Editor, Case Institute of Technology, Cleveland 6, Ohio. Unless otherwise stated, submission of a manuscript is a representation that it has been neither copyrighted, published, nor submitted for publication elsewhere. The authors' institutions are requested to pay a **publication charge** of $8 per page which, if honored, entitles them to 100 reprints without further charge. Instructions will be sent with galley proofs.

Subscriptions, renewals, and orders for back numbers should be addressed to the American Institute of Physics, 57 East 55 Street, New York 22, New York.

Subscription Price	U.S. and Canada	Elsewhere
To members of the American Institute of Physics and Affiliated Societies	$10.00	$11.00
To all others	12.00	13.00

Back Number Prices

Yearly back number rate when complete year is available: $14.00. Single copies: $1.25 each.

Changes of address should be sent to the American Institute of Physics; **corrected proofs** should be sent to the Publication Manager, American Institute of Physics. **Advertising rates** supplied on request. Orders, advertising copy, and cuts should be sent to the American Institute of Physics.

Entered as second class matter January 22, 1937, at the Post Office at Lancaster, Pennsylvania, under the Act of March 3, 1879. Accepted for mailing at the special rate of postage provided for in paragraph (d-2), section 34.40, P. L. & R. of 1948, authorized May 2, 1932.

Journal
of
Applied Physics

Volume 25, Number 2 February, 1954

A Transient-Flow Method for Determination of Thermal Constants of Insulating Materials in Bulk

Part I—Theory

J. H. Blackwell
Department of Physics, University of Western Ontario, London, Ontario, Canada
(Received March 2, 1953)

This is the first of two papers concerning an improved transient-flow method for determining the thermal conductivity and diffusivity of insulating materials in bulk. The work was first suggested by the geophysical problem of determining the thermal constants of natural rock *in situ*. A cylindrical "thermal probe," containing heat-source and thermometer, is inserted in the medium and constants deduced from a record of probe temperature *versus* elapsed time; this method has been used before but the work described here is an attempt to eliminate, or evaluate the effect of, physical idealizations inherent in previous applications. The first paper is concerned with development of a new approximate mathematical treatment, using methods of the operational calculus first suggested by S. Goldstein, in 1932. A subsequent paper will deal with the experimental results obtained with the new theory.

1. GENERAL INTRODUCTION

BASIC inadequacies in mathematical analysis and treatment of experimental data are apparent in previous work with the cylindrical-probe transient flow method for measuring thermal conductivity and diffusivity.[1]

These were drawn to the attention of the author when geophysicists wished to apply the method to the determination of the thermal constants of natural rock *in situ*.

Accordingly it was suggested that the theory be completely revised and experimental tests, using the new theory, be performed under laboratory and field conditions.

In brief, transient probe methods for determining thermal constants may be described as follows: a body of known dimensions and thermal constants (the "probe") which contains a source of heat and a thermometer is immersed in the medium whose constants are unknown. With the aid of suitable theoretical relations, these constants are then deduced from a record of "probe" temperature *versus* elapsed time. (It may be noted that a cylindrical shape for the probe is necessary in the geophysical case, so that the instrument may be inserted in diamond-drill holes already in the rock.)

The deficiencies which have existed in some or all of previous uses of the method are as follows:

(1) It is assumed that, to a first approximation, the "probe" may be represented theoretically as a continuous line-source of heat. A correction of dubious validity is then applied to allow for departure from line-source conditions. Reliability of this method of attack decreases as the probe radius increases and in certain applications (of which the geophysical is one) it is impossible to make the probe sufficiently small.

(2) Thermal contact-resistance at the boundary between "probe" and external medium is assumed zero. This is never true and in particular is a serious disadvantage in the geophysical problem; long drill holes are not very straight and in consequence, the "probe" used must be a loose fit in the hole.

(3) Only semi-intuitive estimates are given of the minimum probe-length to ensure radial-flow conditions and these appear to be unduly exaggerated.

(4) Treatment of experimental data seems relatively crude and does not lead to any systematic evaluation of probable errors in the measured constants.

[1] E. M. F. van der Held and F. G. van Drunen, Physica **15**, 865–881 (1949); F. C. Hooper and F. R. Lepper, J. Am. Soc. Heating Ventilating Engrs. **22**, 129 (1950).

The method has, in consequence, been redeveloped both theoretically and experimentally and it is felt that the objections listed have been largely overcome. The magnitudes of errors introduced by certain simplifying physical assumptions have been established and other minor improvements made.

The new theoretical treatment is described below and experimental results obtained will appear in Part II of the paper. A short summary of the method has already been published.[2]

2. THEORETICAL ANALYSIS

(i) Introduction

It became apparent early in the investigation that two approximate solutions of the heat-flow equation for the probe temperature would be required for evaluation of the thermal constants. Previous workers had used one only, valid for large times. However, inclusion of contact resistance at the probe boundary complicates the large-time solution and makes it desirable to determine the contact resistance "independently." A solution subject to the same boundary and initial conditions but valid for small times makes this determination relatively simple.

Different cylindrical geometries and several methods of introducing heat into the probe were considered, as was the choice of a good or poor conductor for the material of the latter. Examination of theoretical and experimental advantages of the different arrangements led finally to a hollow cylindrical metal probe with constant heat input per unit time. A probe of good conductor can be treated to a first approximation as a perfect conductor and corrections to take account of its finite conductivity applied if necessary. These corrections depend on the location of the probe heater and the point of temperature measurement. Mechanical and electrical design factors indicate the suitability of a spiral heating element wound on a groove on the outside of the probe together with temperature measurement at the inner surface of the hollow cylinder at the center of its length.

Calculations were made to determine the minimum probe length for radial-flow theory to be accurate within experimental error, since the basic theory given below assumes radial flow.

Both large and small-time approximate solutions for the probe-temperature were developed following methods suggested by S. Goldstein[3] as techniques of the operational calculus. In what follows they have been adapted to the equivalent Laplace-transformation procedure.

(ii) Determination of the Probe-Temperature Transform

We wish to solve the radial heat flow equation for an infinite region of one material, bounded internally by a hollow circular cylinder of another (perfectly conducting) material, with the initial and boundary conditions:

(a) Zero initial temperature throughout.

(b) Constant heat-input per unit time to the inner cylinder.

(c) Contact-resistance at the boundary between cylinder and external medium.

Modifications to the results of the basic analysis due to finite probe-conductivity are discussed in (v) below. The validity of the radial-flow assumption is investigated in Appendix I. Let

$\theta_1(t)$, M_1, c_1 = temperature, mass/unit length and specific heat of the "probe,"

$\theta_2(\rho,t)$, K_2, h_2^2 = temperature, thermal conductivity, and diffusivity of external medium,

ρ = radial coordinate,

b = external radius of the "probe,"

t = time,

H = "outer conductivity" at the surface $\rho=b$,

Q = heat supplied/unit probe length/unit time,

$Q'=Q/2\pi b$ and $\alpha=M_1c_1/\pi b$.

Then

$$\frac{\partial^2\theta_2}{\partial\rho^2}+\frac{1}{\rho}\frac{\partial\theta_2}{\partial\rho}=\frac{1}{h_2^2}\frac{\partial\theta_2}{\partial t} \quad b<\rho<\infty: \quad t>0, \tag{1}$$

$$\theta_1=\theta_2=0 \quad t=0, \tag{2}$$

$$-K_2\frac{\partial\theta_2}{\partial\rho}=H(\theta_1-\theta_2) \quad \rho=b: \quad t>0,[4] \tag{3}$$

$$-K_2\frac{\partial\theta_2}{\partial\rho}(2\pi b)=Q-M_1c_1\frac{\partial\theta_1}{\partial t} \quad \rho=b: \quad t>0,[5] \tag{4}$$

i.e.,

$$-K_2\frac{\partial\theta_2}{\partial\rho}=Q'-\frac{\alpha}{2}\frac{\partial\theta_1}{\partial t} \quad \rho=b: \quad t>0. \tag{4a}$$

$$\theta_2 \text{ is bounded as } \rho\to\infty. \tag{5}$$

Laplace transformation of differential equation and boundary conditions with respect to time results in the subsidiary equation and boundary conditions,

$$\frac{d^2\Theta_2}{d\rho^2}+\frac{1}{\rho}\frac{d\Theta_2}{d\rho}=q_2^2\Theta_2 \quad b<\rho<\infty, \tag{6}$$

$$-K_2\frac{d\Theta_2}{d\rho}=H(\Theta_1-\Theta_2) \quad \rho=b, \tag{7}$$

$$=\frac{Q'}{p}-\frac{\alpha p}{2}\Theta_1 \quad \rho=b. \tag{8}$$

$$\Theta_2 \text{ is bounded as } \rho\to\infty, \tag{9}$$

where p is the transform variable; $q_2=p^{\frac{1}{2}}/h_2$.

[2] J. H. Blackwell and A. D. Misener, Proc. Phys. Soc. (London) A64, 1132 (1951).

[3] S. Goldstein, Proc. London Math. Soc. 34, 51 (1932).

[4] H. S. Carslaw and J. C. Jaeger, *Conduction of Heat in Solids* (Oxford University Press, London, 1947), p. 18.

[5] Reference 4, page 16.

$\Theta_1(p)$, $\Theta_2(\rho,p)$ are the L transforms of $\theta_1(t)$, $\theta_2(\rho,t)$, respectively.

Solution of Eq. (6) subject to Eqs. (7), (8), and (9) gives

$$\Theta_1(p)=\frac{2bQ'\Delta}{p[b\alpha p\Delta+2K_2K_1(q_2b)]}, \tag{10}$$

$$\Theta_2(\rho,p)=\frac{2bQ'K_0(q_2\rho)}{pq_2b[b\alpha p\Delta+2K_2K_1(q_2b)]}, \tag{11}$$

where

$$\Delta=\frac{1}{q_2b}K_0(q_2b)+\left(\frac{K_2}{bH}\right)K_1(q_2b). \tag{12}$$

$K_0(x)$, $K_1(x)$ are modified Bessel functions of the second kind, and zero and first orders, respectively.

From the Inversion Theorem of the Laplace Transformation we have, then,

$$\theta_1(t)=\frac{1}{2\pi i}\int_{\gamma-i\infty}^{\gamma+i\infty}\Theta_1(p)e^{tp}dp, \tag{13}$$

$$\theta_2(\rho,t)=\frac{1}{2\pi i}\int_{\gamma-i\infty}^{\gamma+i\infty}\Theta_2(\rho,p)e^{tp}dp. \tag{14}$$

Examination of the integrands of Eqs. (13) and (14) shows that each has a single branch point (at the origin). Furthermore, $\Theta_1(p)$ and $\Theta_2(\rho,p)$ satisfy the conditions of Jordan's Lemma[6] and neither has poles within or on the closed contour $ABCDEFGJK$ as the radius R of the circular arc $\to\infty$. Hence the contour of integration $\mathrm{Re}(p)=\gamma$ may in each case be replaced by the standard equivalent contour $CDEFG$[7] (Fig. 1).

Denoting this contour by Br_2 we may thus write

$$\theta_1=\frac{1}{2\pi i}\int_{Br_2}\Theta_1e^{tp}dp, \tag{15}$$

$$\theta_1=\frac{1}{2\pi i}\int_{Br_2}\Theta_2e^{tp}dp. \tag{16}$$

"Exact" evaluation of integrals in Eqs. (15) and (16) as real infinite integrals is straightforward and that for $\theta_1(t)$ is outlined in Appendix II. These solutions are quite unsuitable however, for the present purpose, i.e., deduction of thermal constants from a temperature-time record, and approximate solutions for θ_1 are developed below.

(iii) Large-Time Approximate Solution for the Probe Temperature

To obtain a large-time solution, we follow the method used by Goldstein.[3] The transform Θ_1 is expanded

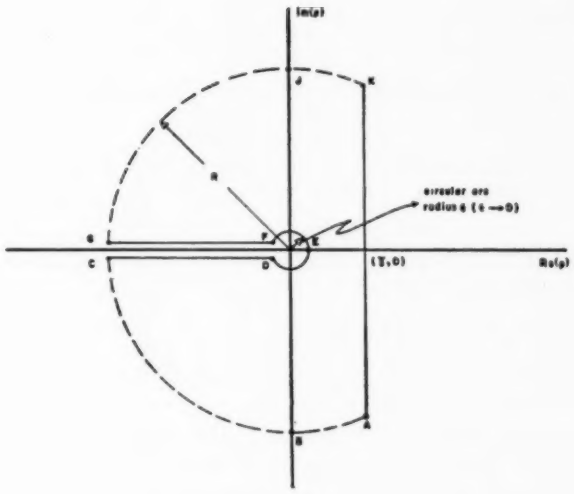


FIG. 1.

formally in ascending powers of p and the integral of Eq. (15) evaluated term-by-term. (Many of the term integrals diverge on the original contour but all converge on Br_2.) The result is a logarithmic term in T ($T=h_2^2t/b^2$) plus an asymptotic expansion in inverse powers of T.

Inserting the ascending-series expansions for the modified Bessel functions in Eqs. (10) and (12) and simplifying, we obtain

$$\Theta_1\simeq-\frac{bQ'}{2K_2}\left[\frac{1}{p}\ln\left(\frac{\beta p}{w}\right)+\frac{b^2}{4h_2^2}\{2\ln(\beta p)-\ln^2(\beta p)-2\}+\frac{b\alpha}{4K_2}\ln^2\left(\frac{\beta p}{w}\right)+0(p)\right], \tag{17}$$

where $\beta=vb^2/4h_2^2$; $\ln w=2K_2/bH$; and $\ln v=2\gamma$. γ is Euler's constant $=0.5722$.

Integrating term-by-term on Br_2, using the results,

$$\frac{1}{2\pi i}\int_{Br_2}ce^{tp}dp=0; \tag{18}$$

$$\frac{1}{2\pi i}\int_{Br_2}\frac{1}{p}\ln(cp)e^{tp}dp=-\left[\ln\left(\frac{t}{c}\right)+\gamma\right]; \tag{19}$$

$$\frac{1}{2\pi i}\int_{Br_2}\ln(cp)e^{tp}dp=-\frac{1}{t}; \tag{20}$$

$$\frac{1}{2\pi i}\int_{Br_2}\ln^2(cp)e^{tp}dp=\frac{2}{t}\left[\ln\left(\frac{t}{c}\right)+\gamma\right]; \tag{21}$$

(where c is a real positive constant and t is positive) we

[6] H. S. Carslaw and J. C. Jaeger, *Operational Methods in Applied Mathematics* (Oxford University Press, London, 1947), second edition, p. 76.

[7] Reference 6, p. 92.

obtain

$$\theta_1(t)\simeq\frac{bQ'}{2K_2}\left[\ln 4T-\gamma+\frac{2K_2}{bH}+\frac{1}{2T}\left\{\ln 4T-\gamma+1-\frac{ah_2^2}{bK_2}\left(\ln 4T-\gamma+\frac{2K_2}{bH}\right)\right\}+0\left(\frac{1}{T^2}\right)\right]. \quad (22)$$

Rigorous mathematical justification of the formal process described above has not yet been made. Nevertheless, comparisons between the results predicted by terms to $0(1/T)$ in Eq. (22) and numerical integration of the exact real integral of Appendix II have been very satisfactory. In addition, these numerical checks (using thermal constants of the order of magnitude of expected unknowns) are useful in setting the minimum value of T for which the expansion Eq. (22) will predict temperature within a given error. It is found, for example, that in typical rock with a probe of suitable size many hours must elapse before the term of $0(1/T)$ can be neglected without introducing appreciable error; on the other hand, if the term of $0(1/T)$ is included, a relatively small minimum T can be tolerated.

An intuitive general justification of the type of approximation process used is suggested by McLachlan.[8]

In the limit $H\to\infty$ (perfect contact), $\alpha\to 0$ (probe walls removed), the solution Eqs. (22) reduces to that of a simpler problem already in the literature.[9]

(iv) Small-Time Approximate Solution for the Probe Temperature

To obtain a small-time solution, another method suggested by Goldstein in the same paper is followed.

An asymptotic expansion of the transform in inverse powers of "p" is made, and this expansion integrated term by term on the original contour $\mathrm{Re}(p)=\gamma$. In the special case considered here, the final result is a series in ascending integral and half-integral powers of "t," but in the general case it is usually a series of repeated integrals of the complementary error function.

Inserting the asymptotic expansions for the modified Bessel functions in Eqs. (10) and (12) and simplifying, we obtain

$$\Theta_1\simeq\frac{2Q'}{\alpha}\left[\frac{1}{p^2}-\left(\frac{2H}{\alpha}\right)\frac{1}{p^3}+\left(\frac{2H^2h_2}{\alpha K_2}\right)\frac{1}{p^{7/2}}+0\left(\frac{1}{p^4}\right)\right]. \quad (23)$$

Making a term by term inverse transformation,

$$\theta_1(t)\simeq\frac{2Q'}{\alpha}\left[t-\frac{H}{\alpha}t^2+\frac{16H^2h_2}{15(\sqrt{\pi})\alpha K_2}t^{5/2}+0(t^3)\right]. \quad (24)$$

Unlike the previous case rigorous justification of this type of approximation is not difficult, e.g., see Carslaw and Jaeger.[10] In dimensionless form the general condition for validity is the converse of the previous case, i.e., $T\equiv h_2^2t/b^2\ll 1$.

(v) Modifications Resulting from Finite Probe Conductivity

In order to investigate the effects of finite probe conductivity, the problem has been solved again as far as the temperature transform, without using the perfect-conductor assumption.

The new solution is, of course, more complex. The region of solution is now a "composite solid," there is a radial temperature gradient in the probe wall and the relative radial locations of heat-injection and temperature measurement become important. Let

K_1, h_1^2 = the conductivity and diffusivity of the probe material;
a, b = the internal and external radii of the hollow probe;
Q = heat supplied/unit probe length/unit time at the outer surface of the probe, $\rho=b$;
Q' = $Q/2\pi b$ (as before);
$\theta_1(a,t)$, $\Theta_1(a,p)$ = the temperature, and temperature transform, respectively at the location of temperature measurement, $\rho=a$;
q_1 = $p^{\frac{1}{2}}/h_1$

and other symbols as before. Then

$$\Theta_1(a,p)=\frac{Q'}{pq_1a}\left[\frac{K_2q_2}{H}K_1(q_2b)+K_0(q_2b)\right]\Big/\Delta', \quad (25)$$

where

$$\Delta'=K_2q_2K_1(q_2b)\{I_1(q_1a)K_0(q_1b)+K_1(q_1a)I_0(q_1b)\}-K_1q_1\left\{K_0(q_2b)+\frac{K_2q_2}{H}K_1(q_2b)\right\}\times\{I_1(q_1a)K_1(q_1b)-K_1(q_1a)I_1(q_1b)\}.$$

We now consider the approximate inverse transform $\theta_1(a,t)$ for large and small times.

Large Times

As before, we expand $\Theta_1(a,p)$ in an ascending series in "p" and integrate term-by-term on contour Br_2. We obtain

$$\theta_1(a,t)\simeq\frac{bQ'}{2K_2}\left[\ln 4T-\gamma+\frac{2K_2}{bH}+\frac{1}{2T}\times\left\{\ln 4T-\gamma+1-\frac{ah_2^2}{bK_2}\left(\ln 4T-\gamma+\frac{2K_2}{bH}\right)-\frac{2h_2^2}{b^2}(\Delta_1+\Delta_2)\right\}+0\left(\frac{1}{T^2}\right)\right], \quad (26)$$

[8] N. W. McLachlan, *Complex Variable and Operational Calculus* (Cambridge University Press, Cambridge, 1946), p. 100.
[9] Reference 4, p. 283.
[10] Reference 6, p. 278.

where

$$\Delta_1=\frac{1}{h_1^2}\left[\tfrac{1}{8}b^2-\tfrac{3}{8}a^2+\tfrac{1}{2}\ln\left(\frac{b}{a}\right)\left\{\frac{a^4}{b^2-a^2}\right\}\right];$$

$$\Delta_2=\frac{1}{h_1^2}\left[\tfrac{1}{8}b^2+\tfrac{1}{8}a^2-\tfrac{1}{2}\ln\left(\frac{b}{a}\right)\left\{\frac{a^2b^2}{b^2-a^2}\right\}\right].$$

It is seen that to $0(1/T)$ this equation is identical with Eq. (22) except for the term $-2h_2^2(\Delta_1+\Delta_2)/b^2$ within the coefficient of $1/T$. This term is usually insignificant in magnitude—it is approximately equal to $-2h_2^2(b-a)^2/3h_1^2b^2$ and, in any event, the curve-fitting procedure for obtaining K_2, h_2^2 which is described below is independent of its value.

Small Times

A "small-time" solution of the ordinary type will be of little use for the composite region under consideration. Since the conductivity of the probe is high, conduction in the probe will pass over into the "large-time" approximation region a few seconds after $t=0$ and it is not experimentally practicable to employ such a short time interval.

Once again Goldstein's paper suggests a suitable alternative method. A "small-time" solution for the poorly conducting external medium is combined with a "large-time" solution for the highly conducting probe. The solution should be useful for calculation provided $b^2/h_1^2\ll t\ll b^2/h_2^2$ i.e., the time region in the immediate vicinity of $t=0$ must be excluded.

To obtain this solution terms in $\Theta_1(a,p)$ involving q_1a and q_1b are expanded in *ascending* series, terms involving q_2b in *descending* series. The series are then combined discarding terms of order higher than $q_1^2a^2$, $q_1^2b^2$, $1/q_2b$ and cross products of these small quantities. Term-by-term inversion is then performed. In the particular case under consideration we obtain

$$\theta_1(a,t)\simeq\frac{2Q'}{\alpha}\left[-\Delta_2+\left\{1-\frac{2(\Delta_1-\Delta_2)H}{\alpha}\right\}t-\frac{H}{\alpha}t^2+\frac{16H^2h_2}{15(\sqrt{\pi})\alpha K_2}t^{5/2}+0(t^3)\right]. \quad (27)$$

If $b-a\ll b$, and $\Delta_1\approx\Delta_2\approx(b-a)^2/6h_1^2$, Eq. (27) becomes

$$\theta_1(a,t)\simeq\frac{2Q'}{\alpha}\left[-\Delta_2+t-\frac{H}{\alpha}t^2+\frac{16H^2h_2}{15(\sqrt{\pi})\alpha K_2}t^{5/2}+0(t^3)\right]. \quad (27a)$$

As was to be expected from physical considerations, the small-time solution is more sensitive to the value of probe conductivity than the large-time solution. The dominant effect for small times is to reduce the temperature at a given time t by a constant amount which is inversely proportional to the probe conductivity.

3. APPLICATION OF THEORETICAL RESULTS

The application of Eqs. (22) and (27) (or 27a) to the determination of thermal constants is straightforward:

Rewrite the "large-time" relation, Eq. (22), in the form,

$$\theta_1(t)\approx A\ln(t)+B+\frac{1}{t}\{C\ln(t)+D\}, \quad (22a)$$

where

$$A=\frac{bQ'}{2K_2}$$

$$B=\frac{bQ'}{2K_2}\left[\ln h_2^2-2\ln b+\ln 4-\gamma+\frac{2K_2}{bH}\right].$$

(Similarly C, D, may be expressed in terms of the constants of the problem but are not required in what follows.)

It can be seen immediately that a fit of suitable experimental data to this expression will yield a value for K_2 directly from the constant A. It is assumed, of course, that the input power to the probe and the probe parameters are known.

In order to evaluate the constant h_2^2 from the value of B, however, it is necessary either that the parameter K_2/bH be negligibly small, i.e., H large (the case of very good thermal contact) or that the value of H be known.

Rewrite the "small-time" relation (27a) in the form

$$Y=\frac{\alpha}{t^2}\left(t-\Delta_2-\frac{\alpha\theta_1}{2Q'}\right)\approx H-\frac{16H^2h_2}{15(\sqrt{\pi})K_2}t^{\frac{1}{2}}. \quad (27b)$$

Then if the quantity y is plotted against $t^{\frac{1}{2}}$ for small t, the parameter H can be read off as the intercept on the y axis. If H is not very small, it may be necessary to include an additional term or terms on the right-hand side of (27b) and fit the solution to the data by a "selected point" method. In this connection it should be borne in mind that the larger the quantity H, the smaller its effect on the determination of K_2, h_2^2. Hence when H is not very small, only a rough value of the parameter is required.

With H known, h_2^2 can be evaluated at once from the constant B.

The final accuracy of determination of K_2, h_2^2 depends intimately on the accuracy of data-fitting to the "large-time" relation. In general, previous workers have neglected the term of $0(1/t)$ in their temperature expansion and determined thermal constants graphically by fitting a straight line by eye to the plot of θ_1 against $\ln(t)$.

As discussed in 2(iii) above, assumption of a straight-line relationship can introduce significant errors in K_2, h_2^2 unless the minimum value of $T(T\equiv h_2^2t/b^2)$ used is high and the contribution of the $1/t$ term therefore

negligible; such a high value of T is usually undesirable from practical considerations.

The method of "least-squares" has been used by the author to fit the experimental data to the *complete* Eq. (22a). A much smaller value of $T_{\min}$ can then be used, the objection to fitting a curve by eye is overcome, and reliable probable errors for the thermal constants can be evaluated (a noticeable omission in previous work on the probe method).

4. ACKNOWLEDGMENTS

The writer wishes to acknowledge with gratitude the financial assistance accorded to him by the Research Council of Ontario during the work described above.

Sincere thanks are recorded also for the facilities of the Physics Department, University of Western Ontario, and for the constant advice and encouragement of the writer's Ph.D. Supervisor, and Department Head, Professor A. D. Misener, and the Research Professor of Physics, Professor R. C. Dearle.

Acknowledgments would be incomplete without recognition also of the helpful comments of the author's teaching colleagues, Dr. E. Brannen, Dr. R. W. Nicholls, and Dr. R. J. Uffen, of whom the latter first suggested the problem.

APPENDIX I

The Assumption of Radial Flow

It is obvious physically that, if the probe is made sufficiently long and temperature measured at the center of its length, the error caused by assuming radial flow can be made as small as we please.

It is necessary, however, to determine a criterion for radial flow conditions within a certain maximum error and this entails at least partial solution of the combined radial/axial flow problem.

The latter is complicated by two considerations:

Firstly, in the case of the present problem, i.e., the probe inserted in a long hole, it is difficult to specify (even approximately) the boundary conditions outside the probe length.

Secondly, analytical complexity is greatly increased by the presence of a layer of good conductor on the inside of the cylindrical hole (the walls of the probe), and by the presence of contact resistance at the probe/medium interface.

It is considered that the presence of a small amount of contact-resistance has no basic influence on the criterion we are trying to determine and it is neglected throughout the analysis outlined below.

This analysis involves the solution of two problems:

(i) We assume the probe walls are of vanishingly small thickness but impose axial boundary conditions which are certainly more severe than those encountered in practice. The temperature at the center of the probe length is then compared with the corresponding radial-flow temperature.

(ii) Using the same axial boundary conditions in the external medium, we attempt to evaluate the effect on the preceding results of the thermal capacity of a finite wall thickness of good conductor. Two simplifications are introduced here:

(a) For convenience, the cylindrical configuration is replaced by an analogous rectangular configuration.

(b) The probe-material is assumed to have infinite conductivity in the "radial" direction, zero conductivity in the "axial" direction. The latter drastic assumption is made necessary by difficulties with boundary conditions; it is justified, in part, by the fact that a very small proportion of the heat injected into the probe is lost from the ends which, in practice, are covered with insulating caps. The necessity for it, however, is the least satisfactory part of the present axial-flow analysis and the writer has begun a new treatment using a prolate spheroidal model for the probe.

The regions and boundary conditions for which the Heat-Flow Equation is solved in the two cases may be specified as follows:

Case (i) The semi-infinite solid bounded by the surface $\rho=b$ internally, and by the planes $z=\pm L$.

The temperature is maintained at zero on $z=\pm L$, $\rho>b$ and there is constant heat flux across the boundary $\rho=b$, $-L<z<L$.

Case (ii) The semi-infinite solid bounded by the planes $z=\pm L$ and a slab of "good conductor," $-d<x<0$. This slab has infinite conductivity in the x direction and zero conductivity in the z direction. There is constant heat input to the slab per unit time and all heat output from the slab is to the region $x>0$, $-L<z<L$. The planes $z=\pm L$ are maintained at temperature zero, $x>0$. Space does not permit of the development of the solutions here[11] but the results and deductions from them are summarized as follows: Case (i) Let θ be the temperature at the inner surface of the cylindrical hole at the center of its length, i.e., at $\rho=b$, $z=0$. Then

$$\theta=\frac{16bQ}{\pi^3K_2}\int_0^T\left[\int_0^\infty \frac{\exp(-uy^2)dy}{y[J_1^2(y)+N_1^2(y)]} \times\sum_{k=0}^{\infty}\frac{(-1)^k\exp\left\{-\dfrac{(2k+1)^2\pi^2b^2u}{4L^2}\right\}}{2k+1}\right]du, \quad (28)$$

where T (as before)$\equiv h_2^2t/b^2$ and Q is flux of heat across the boundary $\rho=b$.

[11] For Case (i) see J. H. Blackwell, Can. J. Phys. **31**, 472–479 (1953).

As $L\to\infty$ the radial/axial flow solution, Eq. (28) reduces to the radial flow solution, i.e.,

$$\theta=\frac{4bQ}{\pi^2K_2}\int_0^T\int_0^\infty\frac{\exp(-uy^2)dydu}{y[J_1^2(y)+N_1^2(y)]};$$
$$=\frac{4bQ}{\pi^2K_2}\int_0^\infty\frac{\{1-\exp(-Ty^2)\}du}{y^3[J_1^2(y)+N_1^2(y)]}. \tag{29}$$

(The latter result is a special case of one published previously.[9]) If now the value of

$$M\equiv\frac{4}{\pi}\sum_{k=0}^{\infty}\frac{(-1)^k\exp\left\{-\dfrac{(2k+1)^2\pi^2b^2T}{4L^2}\right\}}{2k+1}$$

is not appreciably different from unity

$$[\text{i.e., } (\pi^2b^2T/4L^2)\ll 1];$$

then we have, roughly,

$$\theta\approx\frac{1+M}{2}\times\text{radial flow solution.}$$

Let us require the actual temperature to differ from the radial flow solution by less than $1/n$. Then

$$M>1-(2/n).$$

For a maximum axial-flow error of the order of $\frac{1}{2}$ percent, say, we have

$$M>0.99.$$

Values of the sum of the series M are tabulated (e.g., Ingersoll, Zobell and Ingersoll)[12] and for $M>0.99$, we find

$$b^2T/4L^2<0.0158,$$

i.e.,

$$L>(h_2^2t/0.0632)^{\frac{1}{2}}. \tag{30}$$

In using this criterion to determine a minimum length, a value for h_2^2 should be used which is a safe upper limit for the type of material being measured (e.g., 0.015 cgs units for rock) and "t" should be of the order-of-magnitude of the largest time likely to be used in an experiment. It will be noted that the latter will depend both on h_2^2 and the *radial* dimensions of the probe; hence there is some justification for stating the criterion in terms of a length/diameter ratio, e.g.,

$$L/b>(T/0.0632)^{\frac{1}{2}}.$$

[12] Ingersoll, Zobell, and Ingersoll, *Heat Conduction* (McGraw-Hill, Book Company, Inc., New York, 1948), p. 255.

Case (ii) Let θ be the temperature at the boundary between slab and semi-infinite medium at the center of its length, i.e., at $x=0$, $z=0$. Then

$$\theta=\frac{4Q}{\pi K_2}\sum_{k=0}^{\infty}\frac{(-1)^k}{2k+1}\Bigg[\frac{\operatorname{erf}(\beta_kh_2\sqrt{t})}{\beta_k}+\frac{\exp(-\beta_k^2h_2^2t)}{2(A^2+\beta_k^2)^{\frac{1}{2}}}$$
$$\times\{\exp\{[A+(A^2+\beta_k^2)^{\frac{1}{2}}]^2h_2^2t\}$$
$$\times\operatorname{erfc}\{[A+(A^2+\beta_k^2)^{\frac{1}{2}}]h_2\sqrt{t}\}$$
$$-\exp\{[A-(A^2+\beta_k^2)^{\frac{1}{2}}]^2h_2^2t\}$$
$$\times\operatorname{erfc}\{[A-(A^2+\beta_k^2)^{\frac{1}{2}}]h_2\sqrt{t}\}\Bigg], \tag{31}$$

where

$$\beta_k=(2k+1)\pi/2L,$$
$$A=K_2/2Mh_2^2,$$
$$M=\delta_1c_1d.$$

δ_1, c_1, d are the density, specific heat, and thickness, respectively, of the slab, and the remaining symbols have their usual meanings.

(The assumption of zero conductivity of the slab in the z direction permits of the application at $x=0$ of a boundary condition similar in form to the "perfect conductor" boundary condition used in 2 (ii)).

The corresponding linear-flow solution[13] is

$$\theta=\frac{Q}{K_2}[2h_2\sqrt{(t/\pi)}+1/(2A)$$
$$\times\{\exp(4A^2h_2^2)\operatorname{erfc}(2Ah_2\sqrt{t})-1\}]. \tag{32}$$

These results are applied as follows: using values of d, δ_1, and c_1 appropriate to the cylindrical probe being considered and a value of L given by the criterion of Case (i), the values of θ given by Eqs. (31) and (32) are compared for the largest value of t to be used in an experiment.

Should the relative axial-flow error then exceed the maximum originally imposed, the value of L would be increased and the comparison process repeated. In the cases so far examined in this way this has not proved necessary. The results, however, must be treated with a certain amount of circumspection because of the rather artificial boundary condition employed; the problem of the effect of the wall on axial flow error is not yet completely resolved.

It is considered that replacement of the cylindrical problem by an analogous rectangular problem is itself a valid assumption; for example, the reasoning of Case (i), when applied to its rectangular analogue obviously yields the same criterion.

[13] Reference 4, p. 248.

APPENDIX II

The Probe Temperature as a Real Infinite Integral

By the Inversion Theorem of the Laplace Transformation, the probe temperature

$$\theta_1(t)=\frac{1}{2\pi i}\int_{\gamma-i\infty}^{\gamma+i\infty}\Theta_1(p)e^{tp}dp, \tag{13}$$

where $\Theta_1(p)$ is the probe temperature transform given in Eqs. (10) and (12).

To evaluate $\theta_1(t)$ as a real infinite integral a standard artifice is employed. Let

$$\Theta_1(p)=F(p)/p.$$

Then by a well-known theorem of the Laplace Transformation

$$\theta_1(t)=\int_0^t\left[\frac{1}{2\pi i}\int_{\gamma-i\infty}^{\gamma+i\infty}F(p)e^{tp}dp\right]dt. \tag{33}$$

Examination of $F(p)$ shows that like $\Theta_1(p)$ it has a single branch point at the origin and that it is possible in this case also to replace the contour of integration $\mathrm{Re}(p)=\gamma$ by the equivalent contour Br_2. Thus Eq. (33) becomes

$$\theta_1(t)=\int_0^t\left[\frac{1}{2\pi i}\int_{Br_2}F(p)e^{tp}dp\right]dt. \tag{34}$$

Referring to the contour Br_2 ($\equiv CDEFG$) illustrated in Fig. 1, the following substitutions are made:

$$\text{Put}\quad \left.\begin{array}{ll} p=\sigma e^{-i\pi} & \text{on the portion } CD\\ =\epsilon e^{-i\varphi} & \text{on the portion } DEF\\ =\sigma e^{i\pi} & \text{on the portion } FG\end{array}\right\}, \tag{35}$$

where σ, ϵ, φ are real and σ, ϵ positive. Then

$$\int_{Br_2}F(p)e^{tp}dp=\int_0^\infty F(\sigma e^{-i\pi})e^{-t\sigma}d\sigma + \operatorname*{Lt}_{\epsilon\to 0}\int_{-\pi}^{\pi}F(\epsilon e^{i\varphi})\exp(t\epsilon e^{i\varphi})i\epsilon e^{i\varphi}d\varphi - \int_0^\infty F(\sigma e^{i\pi})e^{-t\sigma}d\sigma. \tag{36}$$

Inspection of the form of $F(p)$ shows that the second integral vanishes. Thus,

$$\int_{Br_2}F(p)e^{tp}dp=-2i\int_0^\infty Im[F(\sigma e^{i\pi})]e^{-t\sigma}d\sigma. \tag{37}$$

Reduction of Eq. (37) entails replacement of modified Bessel functions of imaginary argument by ordinary Bessel functions of real argument, and simplification. The substitution $x=(1/h_2)\sigma^{\frac{1}{2}}$ is then made for convenience and we obtain

$$\frac{1}{2\pi i}\int_{Br_2}F(p)e^{tp}dp=\frac{16K_2Q'}{\pi^2b\alpha^2h_2^2}\int_0^\infty\frac{\exp(-h_2^2tx^2)dx}{x[P^2+Q^2]}, \tag{38}$$

where

$$P=xJ_0(bx)+K_2J_1(bx)\left\{\frac{1}{H}x^2-\frac{2}{\alpha h_2^2}\right\},$$

$$Q=xN_0(bx)+K_2N_1(bx)\left\{\frac{1}{H}x^2-\frac{2}{\alpha h_2^2}\right\}.$$

Finally, integrating from 0 to t,

$$\theta_1(t)=\frac{16K_2Q'}{\pi^2b\alpha^2h_2^4}\int_0^\infty\frac{\{1-\exp(-h_2^2tx^2)\}dx}{x^3[P^2+Q^2]}. \tag{39}$$

It is not possible to obtain this result by the more direct method of making the substitutions (35) in the original integral,

$$\frac{1}{2\pi i}\int_{Br_2}\Theta_1(p)e^{tp}dp,$$

since it is found that the real integrals on the component parts of the contour do not converge in this case.

If we take the limit of Eq. (39) as $H\to\infty$ (perfect contact) and make appropriate changes in symbolism, this solution reduces to a special case of one already published.[14]

[14] Reference 4, p. 284.

Relation between Stress-Relaxation Studies and Dynamic Properties of Polyisobutylene

E. CATSIFF AND A. V. TOBOLSKY
Frick Chemical Laboratories, Princeton University, Princeton, New Jersey

(Received April 22, 1953)

Experimental results on stress-relaxation modulus $E_r(t)$ and complex dynamic modulus $E'(\omega)+iE''(\omega)$ of polyisobutylene are now available in the literature. To cover the entire time scale, the effect of changing the temperature of measurement is assumed to be equivalent to a linear change of the time scale. According to the general theory of linear viscoelastic behavior, the functions E_r, E', and E'' can each be computed from any one of them. This is qualitatively correct. The data on E' in the transition region can be represented by the Gauss Error Integral form of equation used by Bischoff, Catsiff, and Tobolsky for E_r of several polymers. The "wedge-box" distribution function proposed by Tobolsky is also compared with the experimental data and shows a reasonably good agreement with experimental results over the entire time scale.

INTRODUCTION

IN previous articles stress-relaxation data on polyisobutylene were collated and a master curve $E_r(t)$ at 25°C was synthesized from data obtained over a wide range of temperatures.[1–5] An idealized distribution of relaxation times $H(\log\tau)$ was presented[5] which, in graphical form, consisted of a wedge and a box.* From $H(\log\tau)$ it was possible to derive analytical expressions for $E_r(t)$ and also for the complex dynamic modulus $E'(\omega)+iE''(\omega)$. The mathematical expressions for $E_r(t)$ gave very close agreement with the experimental master curve. The computed values of $E'(\omega)$ were tabulated to compare with forthcoming experimental data.

Through a cooperative study of the dynamic properties of a NBS (U.S. National Bureau of Standards) sample of polyisobutylene, values of $E'(\omega)$ and $E''(\omega)$ are now available.[6,6a]

In this paper the interrelations between the experimental values of $E_r(t)$, $E'(\omega)$, and $E''(\omega)$ are explored, using the wedge-box idealized distribution function as well as other recently developed mathematical methods.

RELATIONS BETWEEN STRESS-RELAXATION DATA AND DYNAMIC PROPERTIES

If a material is linearly viscoelastic, its properties are completely specified by the distributions of relaxation times $D(\tau)$ or $H(\log\tau)$ which are related as follows:

$$D(\tau)d\tau=H(\log\tau)d\log\tau. \tag{1}$$

The significant relations among $E_r(t)$, $E'(\omega)$, $E''(\omega)$, and $H(\log\tau)$ are expressed by the following exact equations.

$$E_r(t)=\int_{-\infty}^{\infty}H(\log\tau)e^{-t/\tau}d\log\tau, \tag{2}$$

$$E'(\omega)=\int_{-\infty}^{\infty}H(\log\tau)\frac{\omega^2\tau^2}{1+\omega^2\tau^2}d\log\tau, \tag{3}$$

$$E''(\omega)=\int_{-\infty}^{\infty}H(\log\tau)\frac{\omega\tau}{1+\omega^2\tau^2}d\log\tau, \tag{4}$$

$$\Delta E(\omega)=E'(\omega)-[E_r(t)]_{t=1/\omega}=\int_{-\infty}^{\infty}H(\log\tau)f(\zeta)d\log\tau, \tag{5}$$

where

$$f(\zeta)=\frac{1}{1+10^{-2\zeta}}-\exp(-10^{-\zeta}),\qquad \zeta=\log\omega\tau. \tag{6}$$

EXACT COMPUTATION OF $E_r(t)$, $E'(\omega)$, AND $E''(\omega)$ FROM THE WEDGE-BOX DISTRIBUTION

The wedge-box distribution for NBS polyisobutylene was given as[5]

$$H(\log\tau)=0 \qquad \tau<\tau_1, \tag{7}$$

$$H(\log\tau)=2.303P\tau^{-\frac{1}{2}} \qquad \tau_1\leq\tau\leq\tau_2, \tag{8}$$

$$H(\log\tau)=0 \qquad \tau_2<\tau<\tau_3, \tag{9}$$

$$H(\log\tau)=2.303E_0 \qquad \tau_3\leq\tau\leq\tau_m, \tag{10}$$

$$H(\log\tau)=0 \qquad \tau>\tau_m, \tag{11}$$

where

$$P=8.91\times10^3,\quad \tau_1=3.16\times10^{-13},\quad \tau_2=3.98\times10^{-6},$$
$$E_0=7.2\times10^5,\quad \tau_3=4.57\times10^{-5},\quad \tau_m=50.1.$$

The values of the parameters are for the composite stress-relaxation curve at 25°C with $E_r(t)$ measured in dynes/cm² and time measured in hours.

The mathematical expression for $E_r(t)$ obtained by inserting Eqs. (7)–(11) into Eq. (2) has already been

[1] Andrews, Hofman-Bang, and Tobolsky, J. Polymer Sci. **3**, 669 (1948); G. M. Brown and A. V. Tobolsky, J. Polymer Sci. **6**, 165 (1950).
[2] Tobolsky, Dunell, and Andrews, Textile Research J. **21**, 404 (1951).
[3] R. D. Andrews and A. V. Tobolsky, J. Polymer Sci. **7**, 221 (1951).
[4] A. V. Tobolsky and J. R. McLoughlin, J. Polymer Sci. **8**, 543 (1952).
[5] A. V. Tobolsky, J. Am. Chem. Soc. **74**, 3786 (1952).
* The wedge and box correspond to the transition region and rubbery region of viscoelastic behavior, respectively. (See reference 4.)
[6] R. S. Marvin, National Bureau of Standards Report 2305 (1953), submitted to Second International Congress on Rheology, Oxford, July 26–31, 1953.
[6a] Fitzgerald, Grandine, and Ferry, J. Appl. Phys. **24**, 650 (1953).

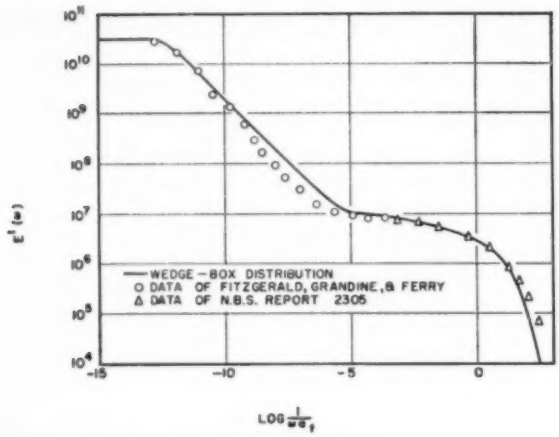


FIG. 1. Dynamic modulus E' of polyisobutylene. Comparison of experimental results (open circles and triangles) with prediction based on "wedge-box" distribution of relaxation times.

given and was shown to agree very well with experimental data.[5]

The analytical expressions obtained by inserting Eqs. (7)–(11) into Eqs. (3) and (4) are

$$E'(\omega)=P(\omega/2)^{\frac{1}{2}}\Big[\tan^{-1}\frac{(2\mu)^{\frac{1}{2}}}{\mu-1}+\tfrac{1}{2}\ln\frac{1+(2\mu)^{\frac{1}{2}}+\mu}{1-(2\mu)^{\frac{1}{2}}+\mu}\Big]_{\mu=\omega\tau_2}^{\mu=\omega\tau_1}+\tfrac{1}{2}E_0\ln[(1+\omega^2\tau_m^2)/(1+\omega^2\tau_3^2)], \quad (12)$$

$$E''(\omega)=P(\omega/2)^{\frac{1}{2}}\Big[\tan^{-1}\frac{(2\mu)^{\frac{1}{2}}}{\mu-1}-\tfrac{1}{2}\ln\frac{1+(2\mu)^{\frac{1}{2}}+\mu}{1-(2\mu)^{\frac{1}{2}}+\mu}\Big]_{\mu=\omega\tau_2}^{\mu=\omega\tau_1}+E_0(\tan^{-1}\omega\tau_m-\tan^{-1}\omega\tau_3). \quad (13)$$

In Figs. 1 and 2 these equations are plotted as solid lines. The experimental points are $3G'(\omega)$ and $3G''(\omega)$ with ω converted to radians per hour. G' and G'' represent modulus measured in shear, and the factor of three is introduced to convert these to Young's moduli E' and E''. The experimental points shown by circles were calculated from recently published data[6a] according to the method of reduced variables, using Ferry's values for J_∞ and a_T.[6b] The experimental points shown by triangles were obtained from the "best" line drawn through the experimental data of several investigators.[6]

It is clear from Fig. 1 that the predicted $E'(\omega)$ is somewhat high at intermediate frequencies but is quite good at the low frequencies which correspond to the experimental time-scale of relaxation at 25°C. The choice of τ_1 made in Eq. (8) was quite arbitrary since the master curve for relaxation did not extend to such very low time-values. It appears from the dynamic modulus data that this arbitrary choice gives excellent agreement with the true asymptotic high-frequency value for $E_r(t)$ and $E'(\omega)$.

The values of $E''(\omega)$ predicted by Eq. (13) also lie above the experimental points in the intermediate frequency range, as is shown in Fig. 2. However, the overall predictions based on the idealized wedge-box distribution must be regarded as quite good. The disparity between experimental dynamic data and the predictions of the "wedge-box" distribution at intermediate frequencies is the result of an unexplained *experimental* discrepancy between relaxation modulus and dynamic modulus in this frequency region, and not to any theoretical shortcoming. This point will be discussed later.

PREDICTIONS BASED ON THE EXPERIMENTALLY MEASURED DYNAMIC MODULUS $E'(\omega)$ OF NBS POLYISOBUTYLENE: THE APPROXIMATE DISTRIBUTION FUNCTION IN THE TRANSITION REGION

Approximation methods have been developed which permit the computation of first and second approximations to $H(\log\tau)$ from any smooth curve which represents $E'(\omega)$, whether this curve is expressed analytically or not. From these approximations to $H(\log\tau)$ one can then predict $E_r(t)$ and $E''(\omega)$ by numerical integration of Eqs. (2) and (4).

Expressing the data for $E'(\omega)$ as an analytical function (although the function may be entirely empirical) is of great convenience for two reasons: (a) in certain cases $H(\log\tau)$ may also be obtained as an analytical function by solution of Eq. (3); (b) if exact solution of $H(\log\tau)$ is impossible, the approximation methods can be applied easily without the inconvenience of graphical determination of slope and curvature.

In this section we will first show that an empirical function exists which expresses the experimental results for $E'(\omega)$ with moderate accuracy. Since an exact solution for $H(\log\tau)$ has not been found, we use approximate

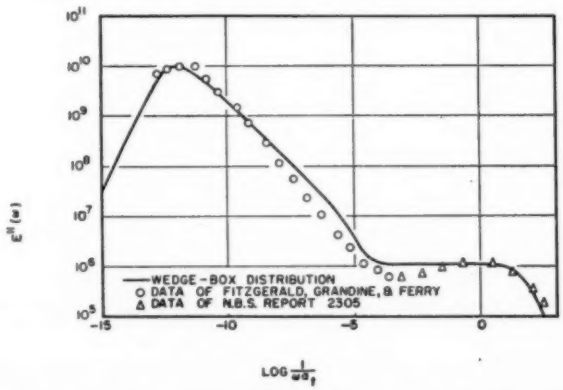


FIG 2. Complex modulus E'' of polyisobutylene. Comparison of experimental results (open circles and triangles) with prediction based on "wedge-box" distribution of relaxation times.

[6b] Ferry, Grandine, and Fitzgerald, J. Appl. Phys. 24, 911 (1953).

expressions $H_1'(\log\tau)$ and $H_{2b}'(\log\tau)$. By inserting these approximations into Eqs. (3) and (4), predicted results for $E'(\omega)$ and $E''(\omega)$ are obtained and compared with experimental results.

The empirical equation which gives a satisfactory fit to experimental values of $E'(\omega)$ at 25° is

$$\log E'(\omega)=8.70+1.82\times\left[2\pi^{-\frac{1}{2}}\int_0^{0.352(\log\omega a_T-9.08)}\exp(-x^2)dx\right]. \quad (14)$$

In Fig. 3, Eq. (14) is compared with experimental data, and the closeness of the fit is shown. Still another test is possible, by plotting on normal probability paper (Fig. 4)

$$100\times\frac{\log E'(\omega)-6.88}{3.64} \textit{ versus } \log 1/\omega a_T. \quad (15)$$

The straight line shown in Fig. 4 is Eq. (14) which is again compared with experimental data. It should be noted that the region in which the experimental values of E' drop below the Gauss Error Integral curve is also the region where an anomalous plateau was found in the plot of loss tangent ($J''/J'=G''/G'=E''/E'$) *versus* reduced frequency. Polyisobutylene appears to be unique in this behavior.[6b]

A first approximation[2,7–9] and a second approximation[10,11] to $H(\log\tau)$ can be obtained from the following

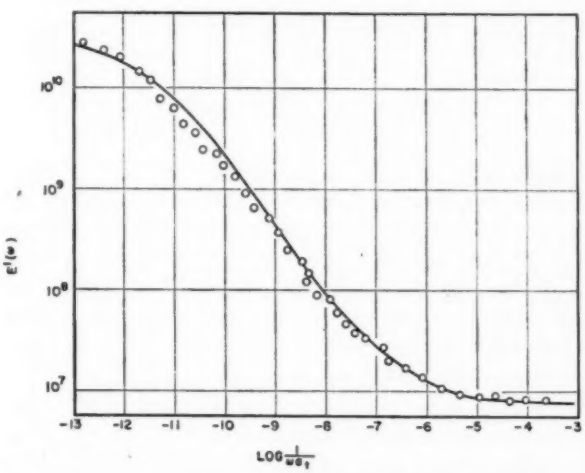


FIG. 3. Dynamic modulus E' of polyisobutylene. Fitting of experimental results (open circles) by Gauss error integral.

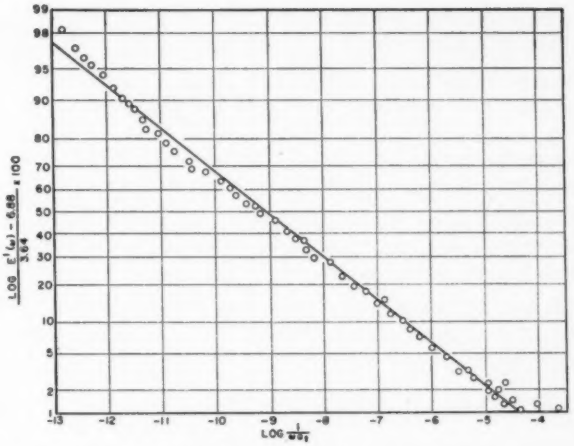


FIG. 4. Dynamic modulus E' of polyisobutylene. Plot of experimental results (open circles) on normal probability paper.

equations:

$$H_1'(\log\tau)=\left[\frac{dE'(\omega)}{d\log\omega}\right]_{\omega=1/\tau}, \quad (16)$$

$$H_{2b}'(\log\tau)=H_1'(\log\tau)/\Gamma\left(1-\frac{m'}{2}\right)\Gamma\left(1+\frac{m'}{2}\right), \quad (17)$$

where

$$m'=-\frac{d\log H_1'(\log\tau)}{d\log\tau}.$$

$H_1'(\log\tau)$ represents the first approximation to $H(\log\tau)$ and $H_{2b}'(\log\tau)$ represents a second approximation.

The results obtained from applying Eqs. (16) and (17) to Eq. (14) are shown in Fig. 5. Also shown are the "wedge" idealization[5] of H and points calculated graphically by Ferry and co-workers.[6b]

In order to check the validity of these approximations, $H_1'(\log\tau)$ and $H_{2b}'(\log\tau)$ were inserted into Eq. (3). Numerical integrations were performed and the results obtained (designated as $E_{1q}'(\omega)$ and $E_{2bq}'(\omega)$, respectively) were compared with $E'(\omega)$. This comparison is made in Fig. 6. If the approximations H_1' and H_{2b}' were perfect, both curves for $E'(\omega)$ shown in Fig. 6 would be identical and would coincide with the experimental data. Though not identical, the curves obtained from H_1' and H_{2b}' give a good fit to the true value of $E'(\omega)$. As was discussed more thoroughly elsewhere, the first approximation, when used to recompute $E'(\omega)$, gives a fairly good over-all fit and is especially good in the high-frequency (low-time) range where the quasistatic glassy modulus, E_1, is being measured.[12] The second approximation gives a very good fit at intermediate and low frequencies, but gives a slightly lower value of the quasistatic glassy modulus in the region of very high frequencies.[13]

[7] T. Alfrey, Jr., *Mechanical Behavior of High Polymers* (Interscience Publishers, Inc., New York, 1948).
[8] Ferry, Sawyer, Browning, and Groth, J. Appl. Phys. **21**, 513 (1950); see also Ivey, Mrowca, and Guth, J. Appl. Phys. **20**, 486 (1949).
[9] R. D. Andrews, Ind. Eng. Chem. **44**, 707 (1952).
[10] Ferry, Fitzgerald, Grandine, and Williams, Ind. Eng. Chem. **44**, 703 (1952).
[11] J. D. Ferry and M. L. Williams, J. Colloid Sci. **7**, 347 (1952).
[12] E. Catsiff and A. V. Tobolsky, report RLT-5 to U. S. Office of Naval Research, Contract No. N6onr-27021, project NR 330-023.
[13] E. Catsiff, report RLT-6 to U. S. Office of Naval Research, Contract No. N6 onr-27021, project NR 330-023.

PREDICTIONS BASED ON THE EXPERIMENTALLY MEASURED DYNAMIC MODULUS $E'(\omega)$ OF NBS POLYISOBUTYLENE: THE PREDICTED STRESS-RELAXATION MODULUS IN THE TRANSITION REGION

The difference function $\Delta E(\omega)$ of Eq. (5) is the most accurate method for predicting stress-relaxation modulus $[E_r(t)]_{t=1/\omega}$, from dynamic modulus $E'(\omega)$ (or *vice versa*), if exact analytical expressions for both of these functions are not known.[14] Numerous approximate relations by means of which $\Delta E(\omega)$ can be obtained from $E'(\omega)$ have been derived.[12,13] The most nearly exact for most real cases is

$$\Delta E_{2b}(\omega)=\Upsilon(m')[H_{2b}'(\log\tau)]_{\tau=1/\omega}, \qquad (18)$$

where

$$\Upsilon(m')=(1/2.303)[(\pi/2)\csc(m'\pi/2)-\Gamma(m')]$$

if $-1<m'<2$ and $m'\neq 0$; for $m'=0$, $\Upsilon(0)=0.2507$.

In Eq. (18), the function $H_{2b}'(\log\tau)$ which appears is defined by Eq. (17). The approximate stress-relaxation modulus $[E_{r,2b}(t)]_{t=1/\omega}$ is defined by the relation

$$[E_{r,2b}(t)]_{t=1/\omega}=E'(\omega)-\Delta E_{2b}(\omega). \qquad (19)$$

Based on experimental values of $E'(\omega)$, this is the best predicted value of $E_r(t)$ which can be obtained without quadrature.

In Fig. 7 the calculated relaxation modulus $E_{r,2b}(t)$ is compared with the experimental stress-relaxation data and also with the dynamic modulus $E'(\omega)$ (Eq. 14) from which $E_{r,2b}(t)$ was derived. The tendency of the experimental stress-relaxation modulus data, $E_r(t)$, to lie above experimental dynamic modulus data, $[3G'(\omega)]_{\omega=1/t}$, in the intermediate frequency range has been pointed out previously.[2,9,14a] This is contrary to the theory of linear viscoelastic behavior as expressed in Eqs. (2)–(5). No final explanation for this anomaly can yet be given, but several alternative explanations will be advanced in a later section.

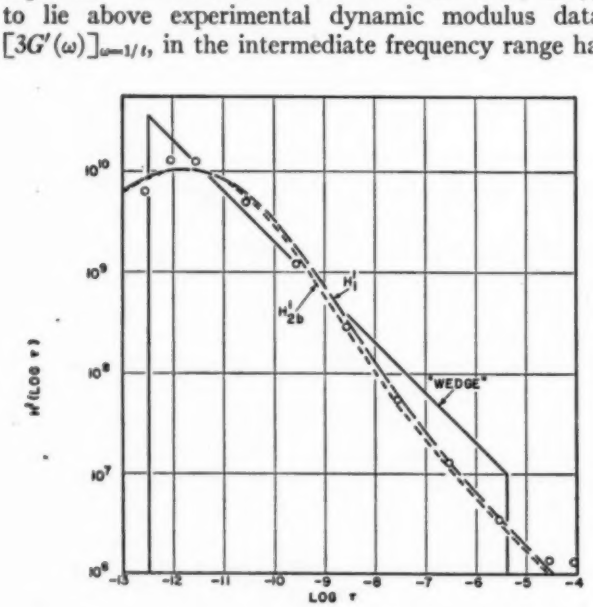


FIG. 5. Approximate distributions of relaxation times for polyisobutylene based on dynamic modulus E'. Open circles show results of graphical determination by Ferry *et al.* (see reference 6b). Comparison with "wedge" distribution.

[14] R. S. Marvin, Phys. Rev. **86**, 644 (1952).

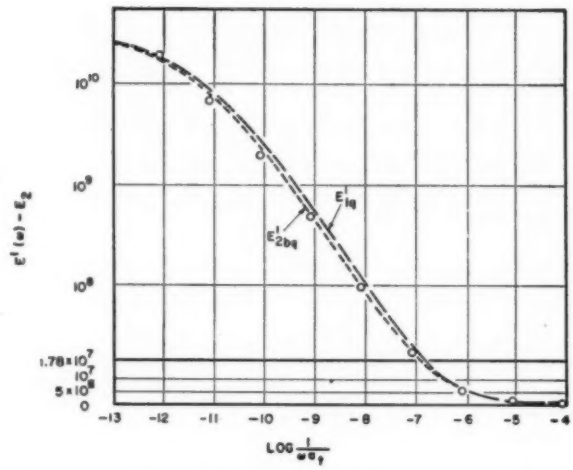


FIG. 6. Dynamic modulus E' of polyisobutylene. Comparison of observed results with results calculated from approximate distributions of relaxation times. (Vertical scale is linear below 1.78×10^{7}.)

The difference between $E_{r,2b}(t)$ and the mathematically exact $E_r(t)$ obtainable in principle from $E'(\omega)$ and Eqs. (2) and (3) is very small.[12,13] It is in fact completely negligible compared with the difference between $E_{r,2b}(t)$ and the experimental relaxation data which are compared in Fig. 7.

PREDICTIONS BASED ON THE EXPERIMENTALLY MEASURED DYNAMIC MODULUS $E'(\omega)$ OF NBS POLYISOBUTYLENE: THE PREDICTED IMAGINARY COMPONENT $E''(\omega)$ OF THE COMPLEX MODULUS IN THE TRANSITION REGION

Simple approximate methods, not requiring quadrature, have been developed[12,13,15] for predicting $E''(\omega)$ from $H(\log\tau)$ and hence from $E_r(t)$ or $E'(\omega)$. A more exact computation of $E''(\omega)$ is obtained by quadrature of Eq. (4) with $H_{2b}'(\log\tau)$ inserted for $H(\log\tau)$.[12,13] We shall designate this approximation $E_{2bq}''(\omega)$. It was shown that $E_{2bq}''(\omega)$ differs inappreciably from $E''(\omega)$ in cases where exact computations can be performed.[13]

$E_{2bq}''(\omega)$, which is the best predicted value of $E''(\omega)$ based on experimental values of $E'(\omega)$, is shown in Fig. 8 where it is compared with the available experimental data for $E''(\omega)$. The general excellence of the fit indicates that the experimental data is of high accuracy.

DISCUSSION

It must first be stated that the results presented here, which have been made possible by the accumulation of

[14a] R. S. Marvin (private communication).

[15] B. A. Dunell and A. V. Tobolsky, Textile Research J. **19**, 631 (1949).

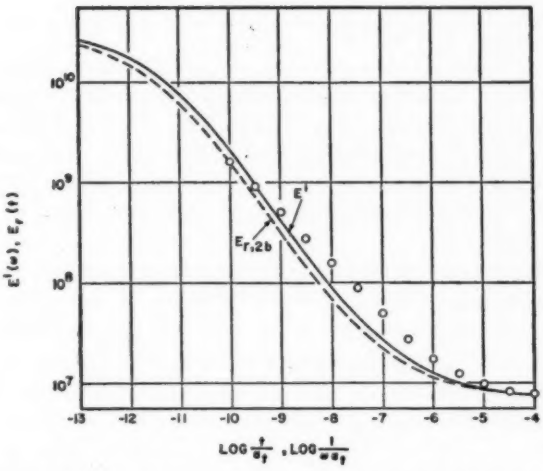


FIG. 7. Predicted stress-relaxation modulus of polyisobutylene, based on the experimental dynamic modulus E'. Open circles show experimental values for stress-relaxation modulus.

stress-relaxation data[1–5] and complex dynamic modulus data,[6,6a] represent the most complete attempt yet made to interpret the behavior of a polymeric substance by means of the general theory of linear viscoelasticity. The results are reasonably gratifying from the point of view of presenting an over-all picture of the viscoelastic properties of polyisobutylene, more complete than has been obtained before. The correlation between stress-relaxation behavior and dynamic properties is qualitatively and semi-quantitatively in accord with the theory of linear viscoelasticity epitomized in Eqs. (2)–(4).

When considered from a more quantitative viewpoint, there is a definite inconsistency in the intermediate frequencies ($5.5<\log\omega a_T<9.5$) between the results for $E_r(t)$ on the one hand and $E'(\omega)$ and $E''(\omega)$ on the other. In this frequency range, the stress-relaxation modulus, $E_r(t)$, is greater than the real part of the dynamic modulus, $E'(\omega)$, although the theory of linear viscoelasticity, and specifically Eq. (5), absolutely forbids this. It should be pointed out that $E_r(t)$ has been determined experimentally in this region; the points shown in Fig. 7 are taken from the composite stress-relaxation curve obtained by the use of the time-temperature superposition principle, not from the curve calculated from the wedge-box idealized distribution function.[5]

It is our belief that the discrepancies are not necessarily due to the inapplicability of Eqs. (2)–(4). It seems more probable that the inconsistencies that have been noted are the result of experimental error in determining $E_r(t)$, $E'(\omega)$, and $E''(\omega)$.

The first experimental procedure that requires scrutiny is the time-temperature superposition principle employed in constructing the master curves for $E_r(t)$, $E'(\omega)$ and $E''(\omega)$.[4,16–18] (The effect of temperature is assumed to be equivalent to a linear change of time scale. This amounts to a horizontal displacement along the log time axis.) It is only by invoking this empirical principle that the complete master curve can be constructed over the entire time scale. In a sense the comparisons made in this paper between predicted values of $E_r(t)$ [derived from the experimental $E'(\omega)$] and measured values of $E_r(t)$ may be regarded as a test of the time-temperature superposition principle, provided that all other sources of experimental error are overlooked. The fact that quantitative correlation is not obtained in the intermediate frequency range indicates that the time-temperature superposition principle is a good approximation but not an absolutely precise principle. This conclusion has already been reached from very careful stress-relaxation measurements.[1] To make stress-relaxation data at different temperatures superpose exactly it was necessary to introduce "stress-factors," as well as "time-factors." Comparison of the temperature dependence of dynamic properties with that of viscosity also leads to this conclusion.[6b]

A second source of error is the necessity for converting stress-relaxation data taken at rather large elongations to $E_r(t)$, and the corresponding necessity to convert dynamic data taken at finite amplitudes to $G'(\omega)$ (the dynamic shear modulus). A further approximation, which is probably very good, is the conversion of shear modulus to Young's modulus, i.e., $E'(\omega)=3G'(\omega)$.

Finally, the experimental difficulties associated with measurement of low-temperature stress-relaxation data and high-frequency dynamic data are considerable. It is perhaps significant that the best correlations between E_r', E', and E'' are obtained at the long-time end of the time scale—both because the experiments are most accurate in this region and also because these data were

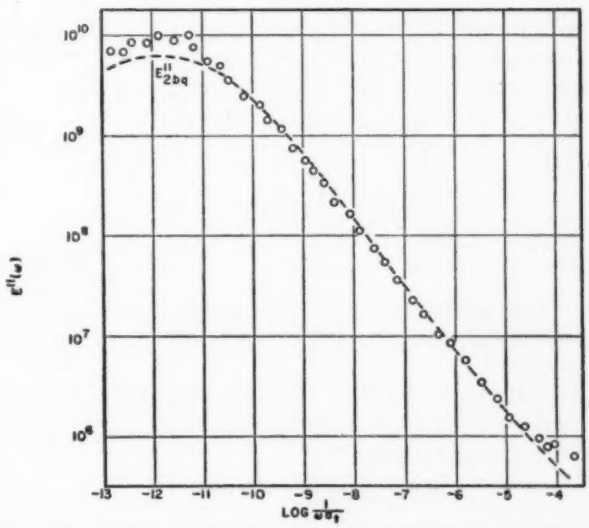


FIG. 8. Complex modulus E'' of polyisobutylene. Comparison of experimental results (open circles) with results calculated from approximate distribution of relaxation times based on dynamic modulus E'.

[16] H. Leaderman, *Creep of Filamentous Materials and other High Polymers* (The Textile Foundation, Washington, D. C., 1943).
[17] A. V. Tobolsky and R. D. Andrews, J. Chem. Phys. 13, 3 (1945).
[18] J. D. Ferry, J. Am. Chem. Soc. 72, 3746 (1950).

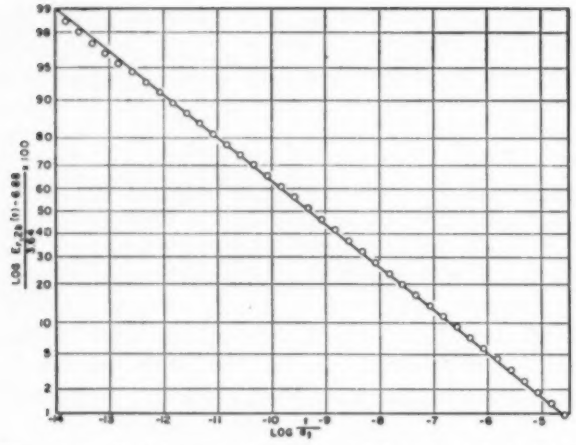


FIG. 9. Predicted stress-relaxation modulus of polyisobutylene, based on $\Delta E_{2b}(\omega)$ (open circles). Fitting of results by Gauss error integral on normal probability paper. Solid line based on Andrews' relation between dynamic and stress-relaxation moduli (see reference 9).

all obtained at 25°C so that the time-temperature superposition principle is not used in this region.

In the light of these several sources of experimental error, a redetermination of $E_r(t)$ of polyisobutylene in the range of reduced frequencies greater than 10^5/hr is highly desirable. This requires operation in the temperature range −50 to −90°C, which poses a not inconsiderable challenge. Another direct line of attack would be the determination of the dynamic properties of polyisobutylene in the same temperature range at such low frequencies as to overlap the stress-relaxation measurements. This probably poses even greater experimental difficulties.

APPENDIX I. THE USE OF THE GAUSS ERROR INTEGRAL FOR REPRESENTATION OF DYNAMIC MODULUS AND STRESS-RELAXATION MODULUS

In a previous publication[19] the stress-relaxation curves of amorphous polymers in the transition region were fitted to the Gauss error integral form of equation

$$E_r(t)=10^{C-B\,\mathrm{erf}.(h\log t_R)}, \tag{20}$$

where

$$\mathrm{erf}x=2\pi^{-\frac{1}{2}}\int_0^x \exp(-u^2)du,$$

$$t_R=t/K.$$

K = temperature-dependent parameter which characterizes viscoelastic behavior as a function of temperature.[6,19]

In this paper we have also proposed to fit the dynamic modulus curves E' of polymers in the transition region by the similar function[19a]

$$E'(\omega)=10^{C'+B'\mathrm{erf}.(h'\log\omega_R)}, \tag{21}$$

[19] Bischoff, Catsiff, and Tobolsky, J. Am. Chem. Soc. **74**, 3378 (1952).

[19a] Recently, Zapas, Shufler, and De Witt, Phys. Rev. **91**, 217 (1953), have also fitted dynamic modulus data with this type of equation.

where

$$\omega_R=K'\omega.$$

From the strict application of the theory of linear viscoelasticity embodied in Eqs. (2)–(4), Eqs. (20) and (21) cannot both be exactly true. However, it is possible for Eqs. (20) and (21) to be approximately valid. Since both equations are empirical, this situation is not unexpected.

In order to test to what extent both equations can be approximately valid and still conform to Eqs. (2) and (3) we have assumed that Eq. (14) was exactly valid for $E'(\omega)$. The already computed function $E_{r,2b}(t)$ [Eq. (19)] represents a very close approximation to the mathematically exact value of $E_r(t)$ which conforms with Eq. (14) and Eqs. (2) and (3). We then attempted to see with what degree of accuracy $E_{r,2b}(t)$ could be represented by Eq. (20). In Fig. 9 the following plot was made on normal probability paper

$$100\times\frac{\log E_{r,2b}(t)-6.88}{3.64}\ \textit{versus}\ \log t. \tag{22}$$

(C' and B' must be equal to C and B since the glassy modulus E_1 and the rubbery modulus E_2 are the same for both static and dynamic measurements.)

The exact values of $E_{r,2b}(t)$ are shown as circles. As seen from this figure these points fall very nearly on a straight line with a slope equivalent to $h'=0.352$. It is clear by comparing expressions (15) and (22) that $C=C'$, $B=B'$ and $h=h'=0.352$.

It has been proposed previously[9] that the curves of $E_r(t)$ *versus* $\log t$ and $E'(\omega)$ *versus* $\log(1/\omega)$ can be virtually superposed by a horizontal shift of 0.2507 log units. Mathematically this means that

$$K'/K=1.781. \tag{23}$$

Another approximation which we attempt to justify in Appendix II gives

$$K'/K=1.622. \tag{24}$$

This would correspond to a horizontal shift of 0.210 log units.

To test Eq. (23), $E'(\omega)$ (on probability paper) *versus* $\log(1/\omega)$ has been shifted over by an amount corresponding to this equation and this shifted curve is compared with $E_{r,2b}(t)$ in Fig. 9. It is clear from this figure that the approximation is reasonably good. It constitutes a very useful approach to the experimental data.

APPENDIX II. CORRELATION OF GAUSS ERROR INTEGRAL REPRESENTATIONS OF DYNAMIC MODULUS AND STRESS-RELAXATION MODULUS

In a previous paper,[12] the first approximation to the difference function when stress-relaxation data is represented by the Gauss error integral was given as

$$\Delta E_1(\omega)=2\gamma\pi^{-\frac{1}{2}}Bh\exp[-(h\log K\omega)^2][E_r(t)]_{t=1/\omega}, \tag{25}$$

where $\gamma=0.5772$, Euler's constant. Hence,

$$\log E_1'(\omega)=\log\{[E_r(t)]_{t=1/\omega}+\Delta E_1(\omega)\}$$
$$=C+B\,\mathrm{erf}(h\log K\omega)$$
$$+\log\{1+2\gamma\pi^{-\frac{1}{2}}Bh\exp[-(h\log K\omega)^2]\}. \quad (26)$$

In every case thus far investigated, the product Bh has been less than unity, so that the second term in brackets is certainly less than unity, and we may expand the logarithm of the bracketed term, giving

$$\log E_1'(\omega)=C+B\,\mathrm{erf}(h\log K\omega)$$
$$+2M\gamma\pi^{-\frac{1}{2}}Bh\exp[-(h\log K\omega)^2]$$
$$-(M/2)(2\gamma)^2\pi^{-1}B^2h^2$$
$$\times\exp[-2(h\log K\omega)^2]+\cdots, \quad (27)$$

where $M=\log e=0.4343$. But, from Eq. (21),

$$\log E'(\omega)=C+B\,\mathrm{erf}(\beta h\log\alpha K\omega), \quad (28)$$

where $h'=\beta h$ and $K'=\alpha K$. If $E'(\omega)$ is set equal to $E_1'(\omega)$,

$$\mathrm{erf}(\beta h\log\alpha K\omega)$$
$$=\mathrm{erf}(h\log K\omega)+2M\gamma\pi^{-\frac{1}{2}}h\exp[-(h\log K\omega)^2]$$
$$-2M\gamma^2\pi^{-1}Bh^2\exp[-2(h\log K\omega)^2]+\cdots, \quad (29)$$

$$2\pi^{-\frac{1}{2}}\int_0^{\beta h\log\alpha K\omega}\exp(-u^2)du$$
$$=2\pi^{-\frac{1}{2}}\int_0^{h\log K\omega}\exp(-u^2)du$$
$$+2M\gamma\pi^{-\frac{1}{2}}h\exp[-(h\log K\omega)^2]$$
$$-2M\gamma^2\pi^{-1}Bh^2\exp[-2(h\log K\omega)^2]+\cdots, \quad (30)$$

$$\int_{h\log K\omega}^{\beta h\log\alpha K\omega}\exp(-u^2)du$$
$$=M\gamma h\exp[-(h\log K\omega)^2]$$
$$-M\gamma^2\pi^{-\frac{1}{2}}Bh^2\exp[-2(h\log K\omega)^2]+\cdots. \quad (31)$$

If β and α are sufficiently close to unity, the integral may be replaced by the product of the integrand and the difference between its limits, so that

$$h\exp[-(h\log K\omega)^2](\beta\log\alpha K\omega-\log K\omega)$$
$$=M\gamma h\exp[-(h\log K\omega)^2]$$
$$-M\gamma^2\pi^{-\frac{1}{2}}Bh^2\exp[-2(h\log K\omega)^2]+\cdots, \quad (32)$$

$$\beta\log\alpha+(\beta-1)\log K\omega$$
$$=M\gamma[1-\gamma\pi^{-\frac{1}{2}}Bh\exp[-(h\log K\omega)^2]+\cdots]. \quad (33)$$

We are primarily interested in the region where $K\omega=1$, so

$$\log\alpha=(M\gamma/\beta)[1-\mu+4\mu^2/3-2\mu^3+16\mu^4/5-\cdots], \quad (34)$$

where $\mu=\gamma\pi^{-\frac{1}{2}}Bh$.

In the simplest case, let $\beta=1(h=h')$ and neglect the terms in μ. Then

$$\log\alpha=M\gamma=0.2507,$$
$$\alpha=e^{\gamma}=1.781. \quad (35)$$

This is the simple relationship between $E_r(t)$ and $E'(\omega)$ which is referred to in Appendix I as Eq. (23) and which has been proposed before.[9] It appears in Fig. 9 as a solid line.

In most cases, while β may be nearly equal to unity, the term $\mu=\gamma\pi^{-\frac{1}{2}}Bh$ is not negligible. Indeed, for the dynamic modulus of NBS polyisobutylene, $\mu=0.209$. If β is set equal to unity, as justified by Fig. 9, and several terms in μ are retained,

$$\log\alpha=0.837M\gamma=0.210,$$
$$\alpha=e^{0.837\gamma}=1.622. \quad (36)$$

This is the second relationship between $E_r(t)$ and $E'(\omega)$, which is given as Eq. (24) in Appendix I. It is not shown in Fig. 9 because it lies but 0.041 units to the right of the line shown.

APPENDIX III

TABLE I. Real component of dynamic modulus of NBS polyisobutylene at 25°C.

	$\log E'(\omega)$, dyne/cm²			
$\log\omega$ (radian/hr)	Computed by Eq. (12)	Reported by Marvin[a]	Found by Ferry *et al*[b]	Computed by Eq. (14)
14.0	10.50	10.56		10.49
13.5	10.50	10.56		10.47
13.0	10.49	10.56		10.43
12.5	10.44	10.54	10.40	10.36
12.0	10.27	10.47	10.27	10.25
11.5	10.04	10.30	10.08	10.10
11.0	9.81	10.03	9.80	9.90
10.5	9.55	9.73	9.51	9.65
10.0	9.30	9.41	9.25	9.34
9.5	9.05	9.08	8.96	9.00
9.0	8.80	8.73	8.63	8.64
8.5	8.55	8.36	8.28	8.29
8.0	8.31	7.98	7.97	7.96
7.5	8.07	7.66	7.67	7.67
7.0	7.81	7.43	7.43	7.43
6.5	7.56	7.23	7.24	7.24
6.0	7.32	7.09	7.10	7.11
5.5	7.12	7.00	7.01	7.02
5.0	7.02	6.95	6.96	6.96
4.5	6.99	6.93	6.93	6.92
4.0	6.97	6.90	6.92	6.90
3.5	6.94	6.87		
3.0	6.89	6.85		
2.5	6.84	6.82		
2.0	6.79	6.79		
1.5	6.73	6.73		
1.0	6.65	6.67		
0.5	6.56	6.58		
0.0	6.45	6.46		
−0.5	6.30	6.29		
−1.0	6.07	6.06		
−1.5	5.66	5.76		
−2.0	4.91	5.35		

[a] See reference 6.
[b] See reference 6a.

TABLE II. Imaginary component of dynamic modulus of NBS polyisobutylene at 25°C.

logω (radian/hr)	log$E''(\omega)$, dyne/cm^2 Computed by Eq. (13)	Reported by Marvin[a]	Found by Ferry *et al*[b]	logω (radian/hr)	log$E''(\omega)$, dyne/cm^2 Computed by Eq. (13)	Reported by Marvin[a]	Found by Ferry *et al*[b]
14.0	8.53	8.20		5.5	6.95	6.49	6.55
13.5	9.02	8.70		5.0	6.56	6.19	6.25
13.0	9.51	9.19		4.5	6.19	5.96	5.99
12.5	9.89	9.66	9.89	4.0	6.07	5.83	5.87
12.0	10.00	9.97	10.03				
				3.5	6.06	5.79	
11.5	9.90	10.12	9.98	3.0	6.05	5.80	
11.0	9.72	10.05	9.82	2.5	6.05	5.85	
10.5	9.51	9.83	9.58	2.0	6.05	5.93	
10.0	9.27	9.57	9.32				
				1.5	6.05	5.99	
9.5	9.03	9.25	9.08	1.0	6.05	6.04	
9.0	8.79	8.90	8.79	0.5	6.05	6.08	
8.5	8.54	8.53	8.47	0.0	6.05	6.09	
8.0	8.29	8.20	8.15				
				−0.5	6.04	6.06	
7.5	8.04	7.86	7.81	−1.0	6.00	5.97	
7.0	7.80	7.52	7.48	−1.5	5.86	5.79	
6.5	7.54	7.17	7.16	−2.0	5.52	5.58	
6.0	7.28	6.85	6.85				

[a] See reference 6.
[b] See reference 6a.

JOURNAL OF APPLIED PHYSICS VOLUME 25, NUMBER 2 FEBRUARY, 1954

Ferrites for Microwave Circuits and Digital Computers

E. ALBERS-SCHOENBERG
General Ceramics and Steatite Corporation, Keasbey, New Jersey
(Received May 25, 1953)

Ferrites of the system MgO-MnO-Fe_2O_3 have been investigated, and two new classes of magnetic materials have been developed. One class of ferrite has low losses and shows the Faraday rotation effect at microwave frequencies, which makes it useful in new types of microwave components. A second class of ferrite has a rectangular hysteresis loop and inherent high resistivity which allows digital computer magnetic memories to be operated at much greater speeds than was possible with the familiar magnetic metal alloys. Both classes of ferrites are now commercially available.

IN an early stage of the investigation of ferromagnetic ferric-oxide compounds, two ferrites with very different magnetic properties attracted attention: magnesium ferrite and manganese ferrite. The first was remarkable for its high insulating qualities (10^9 ohm-centimeter) while the second had an unusually high permeability for a single ferrite. For a material combining magnetic and dielectric properties, magnesium ferrite appeared to be useful; its refractory nature, however, prohibited the commercial preparation of a dense body at readily-available temperatures. Hence manganese oxide was added as a third component to promote the formation of a dense body and at the same time to contribute to the magnetic strength of the material. Further study of this three-oxide system with respect to variations in composition and firing conditions produced a number of ferrites with unexpected properties and new technical applications.

The first commercial ferrite of this system was Ferramic A–34.[1] As a result of low absorption loss, this material proved useful in microwave applications.[2,3] This was an unusual type of ferrite since it contained an excess of magnesium oxide which was not balanced by an equivalent amount of iron oxide; the mole ratio of the composition was MgO:MnO:Fe_2O_3=52:7:41. It is a peculiarity of this oxide system that the range of useful compositions extends far over to the magnesia excess side (see Fig. 1). The location of composition A–34 is shown near the center of the enclosed areas on the diagram. Starting from composition A–34, the investigation went in two directions; each direction yielding a class of ferrites with unique properties. In the upper area of the diagram is located composition MF–1331, which represents a much-improved variation of A–34 for microwave applications. Dielectric and magnetic properties for Ferramic A–34 and MF–1331 are listed in Table I. At one megacycle, the dielectric losses of MF–1331, although large compared with true high-frequency dielectrics, are small when compared with other ferrite bodies. The dielectric constant does not

[1] Snyder, Albers-Schoenberg, and Goldsmith, "Magnetic ferrites," Elec. Mfg. 44, 86 (1949).

[2] D. Polder, Phil. Mag. 40, 99 (1949).
[3] Rado, Wright, and Emerson, Phys. Rev. 80, 273 (1950).

change up to microwave frequencies while the dielectric loss decreases with increasing frequency. The application to microwave circuitry depends on the ability of this ferrite when in a magnetized state to rotate the plane of polarization of an electromagnetic wave which passes through it; the Faraday rotation effect is well known in optics.[4–6] Since in addition the material has low dielectric losses, MF–1331 is finding use in microwave switches, modulators, and phase changers.

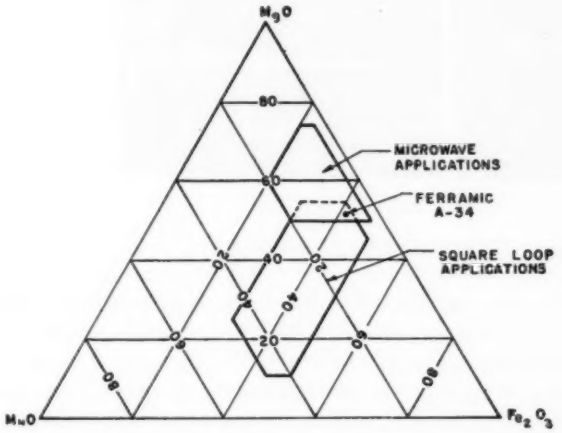


FIG. 1. Triaxial diagram (all figures mole percent) of the system MgO-MnO-Fe_2O_3 showing favorable working areas.

Ferramic A–34 deserved attention for more than its microwave properties. It was the first composition which gave evidence that a square hysteresis loop ferrite could be developed.[1] Since the square-loop metal alloys have limited use in high-frequency circuits, in particular, high speed pulse applications, the investigation of the A–34 took on a new importance.[7] The lower area on the

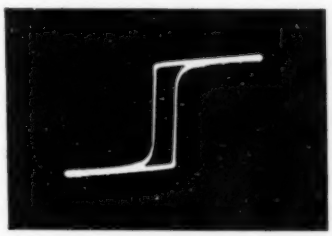


FIG. 2. Saturated hysteresis loop of MF–1326–B.

triaxial diagram (Fig. 1) shows the region which is the source of compositions having well-developed square hysteresis loops. In this region is located composition MF–1326–B, which is now widely used in magnetic memory applications.

The functioning of a ferrite memory core in a digital computer is through magnetic storage of information, due to the high degree of squareness of the hysteresis

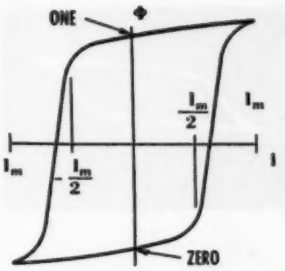


FIG. 3. Characteristic curve of square loop ferrite core explaining the meaning of the "squareness ratio."

loop.[7–14] The features of a good square hysteresis loop are given with reference to Fig. 2 which shows a 60-cps hysteresis loop photographed from an oscilloscope. The important requirements of this loop are as follows: the magnitude of flux density at $H=0$ should be as high as possible so that the Br/Bs ratio approaches unity; the

TABLE I. Table of properties of two ferrites for microwave applications.

		A–34	MF-1331
Initial permeability (1 Mcps)	μ_0	15	50
Maximum permeability (DC)	μ_{max}	100	350
Saturation flux density (DC) gauss	B_s	850	1600
Retentivity (DC) gauss	B_r	615	1300
Coercive force (DC) oersteds	H_c	3.5	2.5
Dielectric constant	ϵ'	10	13
Dielectric power factor			
10^6 cps	tanδ	0.035	0.03
10^{10} cps		0.01	0.003
Resistivity	Ohm-cm	8×10^8	1×10^9

outside corners should be sharp with the inside corners well developed; the loop should have steep sides. These characteristics may be combined into one measurable figure of merit called the squareness ratio. As illustrated in Fig. 3, this is the ratio of magnetic flux at $-\frac{1}{2}I$ max to the magnetic flux at $+I$ max for a particular hysteresis loop. The maximum squareness ratio for MF–

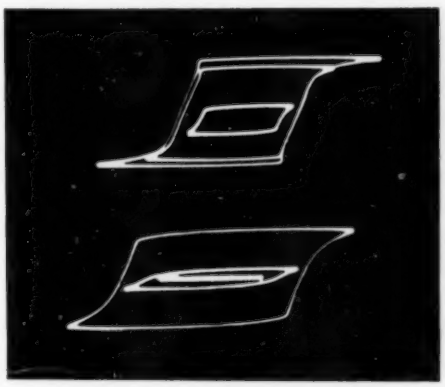


FIG. 4. Family of hysteresis loops of MF–1374 (top) and MF–1326–B (below) at H max=0.3; 0.7; 1.4 oersteds.

[4] N. G. Sakiotis and H. N. Chait, Proc. Inst. Radio Engrs. **41**, No. 1 (1953).
[5] P. J. Allen, Proc. Inst. Radio Engrs. **41**, No. 1 (1953).
[6] C. L. Hogan, Bell System Tech. J. **31**, 1 (1952).
[7] E. Gelbard, Tele-Tech **11**, 50 (May, 1952).
[8] J. W. Forrester, J. Appl. Phys. **22**, 44 (1951).
[9] W. N. Papian, Proc. Inst. Radio Engrs. **40**, 475 (1952).
[10] J. A. Rajchman, RCA Rev. **13**, 183 (1952).
[11] An Wang and Way Dong Woo, J. Appl. Phys. **21**, 49 (1950).
[12] H. Billing, Naturwiss. **40**, Heft 2 (1953).
[13] D. R. Brown and E. Albers-Schoenberg, Electronics **26**, 146 (1953).
[14] E. Albers-Schoenberg Ann. Rep. Comp. El. Insul. **1952**, 51.

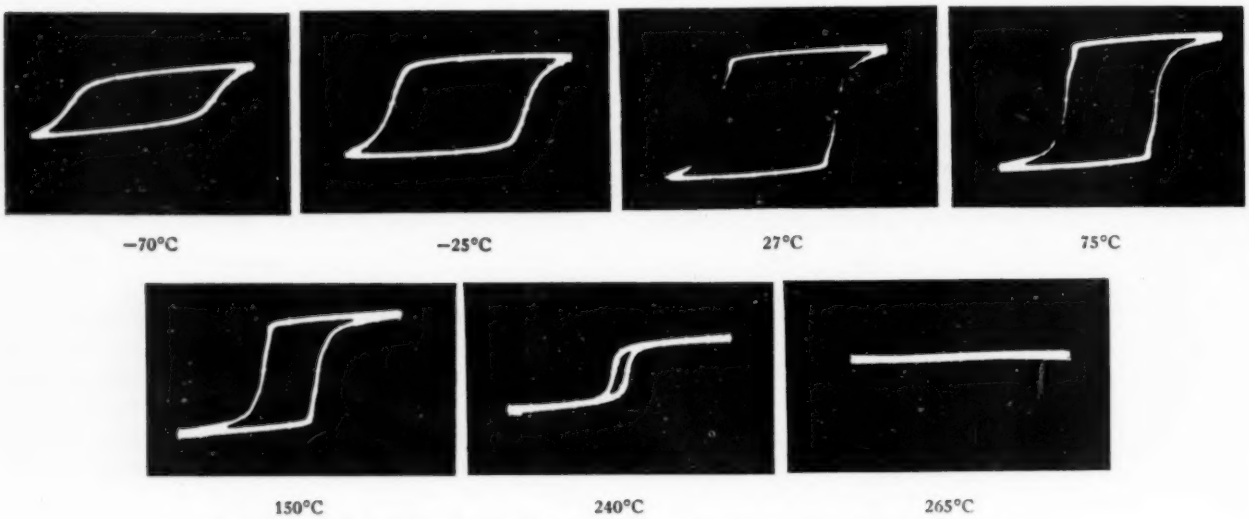


FIG. 5. Hysteresis loops (60 cps-H max = 2 oersteds) *vs* temperature for MF–1326–B.

1326–B is 0.8 and for its related composition MF–1374 is 0.95.

Figure 4 shows a family of hysteresis loops taken at different field strengths for MF–1326–B and MF–1374. The hysteresis loops were photographed at field strengths of 0.3, 0.7, and 1.4 oersteds. Magnetic properties for these two materials are given in Table II.

TABLE II. Table of properties of two square loop ferrites.

		MF-1326-B	MF-1374
Initial permeability (1 Mcps)	μ_0	40	48
Maximum permeability (DC)	μ_{max}	515 (1000 gauss)	1600 (1400 gauss)
Saturation flux density (DC)	B_s	1780	1950
Retentivity (DC) gauss	B_r	1590	1870
Coercive force oersteds	H_c	1.5	0.7
Access time microseconds	t	1–2	4–7
B_r/B_s ratio		0.89	0.96
Maximum squareness ratio $\Phi(-I_m/3)/\Phi(I_m)$	R_s	0.8	0.95

Figure 5 shows the effect of temperature on the shape of the hysteresis loop of MF–1326–B. This loop represents a field strength of 2 oersteds which is approximately equivalent to the field strength at which optimum pulse test results are obtained. Within the range of temperature from −25°C to +75°C, the squareness ratio remains sufficiently high for magnetic memory applications.

When used in a digital computer, a significant property of a square loop core is its switching or access time, which is defined as the time required to drive the core from the state of $-Br$ to the state of $+Bs$, or *vice versa*. Early in the development of the square loop ferrites, these materials were found to have extremely short switching times (as low as 0.5 microsecond) which indicated that they were particularly suited for use as memory units in high-speed electronic digital computers. It is interesting to notice that there appears to be a relationship between coercive force and switching time. For example, the first square loop ferrites had a switching time of less than one microsecond but a coercive force of 2.5 to 3 oersteds; these cores had to be "driven hard" to reverse their state of magnetization. When with further development the coercive force was reduced to 0.5 to 0.7 oersteds, the switching time was simultaneously increased to 4–7 microseconds. In general, a relatively large amount of manganese oxide results in a body having high saturation and long switching time while a large amount of magnesium oxide has the effect of reducing both these properties.

Although square loop ferrites have only recently arrived at the stage of widespread technical usefulness, many thousands of cores are already being used in digital computer memory applications. Increasing interest in square loop and microwave applications demands a fundamental understanding and a further investigation of this oxide system.

JOURNAL OF APPLIED PHYSICS VOLUME 25, NUMBER 2 FEBRUARY, 1954

The Synthesis of X-Ray Spectrometer Line Profiles with Application to Crystallite Size Measurements*

LEROY ALEXANDER
Department of Research in Chemical Physics, Mellon Institute, Pittsburgh, Pennsylvania
(Received July 3, 1953)

An improved convolutional method is used to synthesize theoretical line profiles generated by the joint action of seven instrumental weight functions characteristic of the x-ray spectrometer. These include in addition to the five geometrical factors described in a previous publication: (VI) a so-called misalignment function and (VII) a spectral impurity function. The method is applied to the new spectrometer design incorporating a narrow x-ray source, Soller slits, and an angular range $2\theta=0$ to 165° or higher. Good agreement between theoretical and experimental line profiles is achieved for quartz reflections at $2\theta=24.0$ and 90.7°. The improved treatment is used to define new curves for correcting experimental spectrometer line breadths for instrumental broadening effects. It is demonstrated that spectrometers of the new design are capable of measuring crystallite dimensions as large as 3000A.

I. INTRODUCTION

EARLIER publications[1,2] of the present author dealing with the effect of geometrical factors upon spectrometer line profiles were restricted to the forward-reflection region and did not include a treatment of Soller-slit collimators. The subsequent modification of the design of x-ray spectrometers, incorporating Soller slits and a narrow x-ray source and extending the angular range into the back-reflection region, has necessitated a review and some expansion of the previous treatment. The present investigation makes a more effective use of the convolution principle[2-4] to predict theoretically the nature of the line profiles to be expected under various experimental conditions. A useful result has been the establishing of curves appropriate to the new spectrometer design for the correction of line profiles for geometrical broadening. These curves are of practical value in the determination of crystallite size from line broadening.

As explained in the previous treatment,[2] for powders of sufficiently large crystallite dimensions the observed line profile is determined by the joint action of five specific instrumental factors, or weight functions: (I) x-ray source profile, (II) flat rather than curved sample surface, (III) vertical divergence, (IV) absorption of the incident beam by the specimen, and (V) width of the receiving slit. The new treatment embraces two additional factors, which under certain circumstances exert their effects upon the observed line shape: (VI) the misalignment function and (VII) the spectral impurity function. The misalignment function, which is so designated for lack of a better name, is a nonspecific and empirical function sometimes found to be essential in achieving close agreement between the synthetic and experimental line profiles. It is likely, though not certain, that this factor is the result of a somewhat random combination of small, residual imperfections in the alignment of the several components of the spectrometer. Possibly a specific geometrical aberration of the system is responsible, but, if so, the writer has not yet diagnosed it. The spectral impurity function, although not of instrumental origin, causes significant line broadening at the larger angles and must be reckoned with in matching theoretical and observed profiles.

The present investigation has demonstrated the inaccuracy of a statement made in the introduction of the previous paper,[2] "In the newer models vertical divergence is practically eliminated by the use of Soller slits . . . ." This was an uncritical expression of the popular opinion prevalent at the time of the introduction of the new spectrometer design. The outcome of the present studies corroborates the recent findings of Eastabrook[5] that the vertical-divergence profile generated by Soller-slit collimators is of the same order of breadth as that produced by spectrometers without Soller slits, although the shape of the profile is modified. Vertical-divergence effects with Soller slits are weighty at small Bragg angles but insignificant in comparison with other sources of broadening at large angles.

II. SYNTHESIS OF LINE PROFILES BY THE CONVOLUTION METHOD

As explained previously,[2] if the superposition theorem holds, the line profile $h(\epsilon)$ generated by a spectrometer is the convolution of the pure diffraction function $f(\epsilon)$ and the over-all weight function of the apparatus $g(\epsilon)$,

$$h(\epsilon)=\int_{-\infty}^{+\infty} g(\zeta)f(\epsilon-\zeta)d\zeta, \quad (1)$$

where the variables ϵ and ζ represent angular displacements from the theoretical diffraction angle 2θ, their sign being positive in the direction of increasing 2θ. The function g will be here understood to include the influence of the natural spectral distribution of wavelengths

* Based in part upon material presented before the American Crystallographic Association at Tamiment, Pennsylvania, on June 20, 1952.

[1] L. Alexander, J. Appl. Phys. **19**, 1068 (1948).
[2] L. Alexander, J. Appl. Phys. **21**, 126 (1950).
[3] R. C. Spencer, Phys. Rev. **55**, 239 (1939).
[4] R. C. Spencer, J. Appl. Phys. **20**, 413 (1949).
[5] J. N. Eastabrook, Brit. J. Appl. Phys. **3**, 349 (1952).

in the ideally monochromatic radiation employed.[6] This net instrumental function may also be resolved into the several specific instrumental functions, $g_{\rm I}$, $g_{\rm II}$, etc., equivalent to the respective factors enumerated in the introduction. The function $g(\epsilon)$ can then be derived from the seven specific functions by means of the following six successive convolutional operations, the order in which the several functions are chosen being irrelevant:

$$\left.\begin{aligned}
g_{\rm I,\,II}(\epsilon) &= \int_{-\infty}^{+\infty} g_{\rm I} g_{\rm II}(\epsilon-\zeta)d\zeta,\\
g_{\rm I-III}(\epsilon) &= \int_{-\infty}^{+\infty} g_{\rm I,\,II} g_{\rm III}(\epsilon-\zeta)d\zeta,\\
g_{\rm I-IV}(\epsilon) &= \int_{-\infty}^{+\infty} g_{\rm I-III} g_{\rm IV}(\epsilon-\zeta)d\zeta,\\
g_{\rm I-V}(\epsilon) &= \int_{-\infty}^{+\infty} g_{\rm I-IV} g_{\rm V}(\epsilon-\zeta)d\zeta,\\
g_{\rm I-VI}(\epsilon) &= \int_{-\infty}^{+\infty} g_{\rm I-V} g_{\rm VI}(\epsilon-\zeta)d\zeta,\\
g_{\rm I-VII}(\epsilon) = g(\epsilon) &= \int_{-\infty}^{+\infty} g_{\rm I-VI} g_{\rm VII}(\epsilon-\zeta)d\zeta.
\end{aligned}\right\} \quad (2)$$

Once the over-all instrumental profile $g(\epsilon)$ has been computed for a given set of experimental conditions, the final line profile can be synthesized for a powder specimen of any given crystallite size by performing the convolution (1) of $g(\epsilon)$ and the appropriate pure diffraction line profile $f(\epsilon)$. When the crystallite size exceeds about 1000A at small, and about 3000A at large Bragg angles, the breadth of the $f(\epsilon)$ profile becomes negligible in relation to that of $g(\epsilon)$, and the observed line profile is given simply by $g(\epsilon)$.

The present convolutional synthesis differs in two respects from that previously published.[2] First, the contribution of the pure diffraction function $f(\epsilon)$, if any, is not dealt with until after the development of the net instrumental function $g(\epsilon)$, according to the scheme given by Eqs. (2). Second, greater flexibility and exactness are gained by expressing the successive convolutions, $g_{\rm I,\,II}$, $g_{\rm I-III}$, etc., in this scheme as analytically faithful sets of numerical data rather than by attempting to express them, more or less successfully, as simple mathematical formulas (see, for example, Table II of reference 2). For spectrometers without Soller slits the previously published descriptions of the first five instrumental weight functions[2] are unchanged save for one minor exception. Eastabrook's excellent treatment of vertical-divergence effects[5] demonstrates that the angular limits of the vertical-divergence profile, $g_{\rm III}=|\epsilon|^{-\frac{1}{2}}$, are $\epsilon=0$ and $-(1/4)\delta^2\cot\theta$ rather than $\epsilon=0$ and $-(1/8)\delta^2(1+\cos2\theta)^2\cot2\theta$, as deduced by the present writer using a more approximate method. The two negative limits do not differ greatly at small Bragg angles, where the vertical-divergence effect is relatively large, but they differ increasingly at the higher angles, where, however, it happens that the magnitude of the effect is so small as to resist detection by ordinary experiments.[7]

Table I recapitulates the nature of the five instrumental weight functions previously described[2] for spectrometers without Soller slits. In addition the table includes the necessary expressions for spectrometers with Soller slits and for the extended angular range $2\theta=0$ to 180°. Figure 1 shows the appearance of the six instrumental weight functions as computed for old and new spectrometers operating under typical experimental conditions. These include a diffraction angle of $2\theta=24°$ and a medium linear-absorption coefficient, $\mu=34$. At this rather small angle the asymmetry resulting from vertical divergence is large, whereas the spectral-impurity broadening is so small as to be negligible. It should be noted that functions $g_{\rm I}$, $g_{\rm V}$, and $g_{\rm VI}$ broaden

TABLE I. The seven instrumental weight functions determining x-ray spectrometer line profiles.

	Function	Form of the function	Definitions or remarks[a]
$g_{\rm I}$	Source	$\exp-k_{\rm I}^2\epsilon^2$ (approximate form)	$k_{\rm I}=1.67/w_{\rm I}$
$g_{\rm II}$	Flat sample	$\lvert\epsilon\rvert^{-\frac{1}{2}}$	$\epsilon=0$ to $\epsilon_m=-\dfrac{\gamma^2\cot\theta}{114.6}$ degrees
$g_{\rm III}$	Vertical divergence	No Soller slits: $\lvert\epsilon\rvert^{-\frac{1}{2}}$	$\epsilon=0$ to $\epsilon_m=-\dfrac{\delta^2\cot\theta}{(4\times57.3)}$ degrees
		Soller slits: no simple expression, see reference 5 for profiles	
$g_{\rm IV}$	Absorption	$\exp-k_{\rm IV}\epsilon$ (for thick specimens)	ϵ is always negative, $k_{\rm IV}=4\mu l/114.6\sin2\theta$
$g_{\rm V}$	Receiving slit	Constant	Angular limits $\epsilon-w_s/2$ and $\epsilon+w_s/2$
$g_{\rm VI}$	Misalignment	$1/(1+k_{\rm VI}^2\epsilon^2)$ (in a typical case examined)	$k_{\rm VI}=2/w_{\rm VI}$
$g_{\rm VII}$	Spectral impurity	$1/(1+k_{\rm VII}^2\epsilon^2)$ (approximate form)	$k_{\rm VII}=2/w_{\rm VII}$ and $w_{\rm VII}=0.0424\tan\theta_0$ for Cu$K\alpha$ radiation[b]

[a] The coefficients are such as to express all angles in degrees.
ϵ = angular displacement from the theoretical diffraction angle, $2\theta_0$.
w = angular width of profile at half-maximum intensity.
γ = horizontal divergence of the primary x-ray beam.
δ = vertical divergence of the primary x-ray beam.
l = spectrometer radius in cm.
μ = linear absorption coefficient of the specimen.
w_s = angular width of the receiving slit.
[b] See reference 6.

[6] H. Ekstein and S. Siegel, Acta Cryst. **2**, 99 (1949).

[7] L. Alexander, Brit. J. Appl. Phys. **4**, 92 (1953).

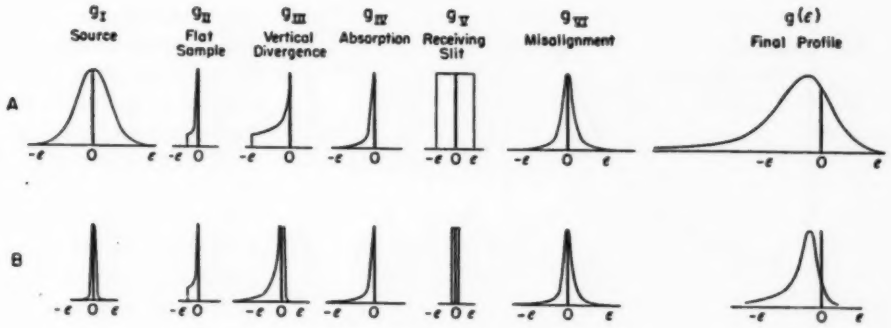


FIG. 1. The six instrumental weight functions for spectrometers of *A* old and *B* new designs computed for typical experimental conditions.

the profile symmetrically, but that g_{II}, g_{III}, and g_{IV} broaden it asymmetrically and displace the peak of the final profile toward lower angles. It can be seen from the figure that the striking improvement in resolving power displayed by spectrometers of the new design springs from the greatly reduced width of the x-ray source when the focal spot is viewed laterally instead of longitudinally and from the attendant use of a narrower receiving slit. In the final section of this paper it is shown that this conspicuous reduction in the line-profile breadth permits a significant increase in the range of crystallite size measurements compared with the potentialities of the older spectrometers or photographic methods in general. With the exception of g_I and g_V, the instrumental weight functions are little changed. The magnitude of the misalignment function given in Fig. 1 was chosen empirically so as to effect optimum agreement between theory and experiment for the particular spectrometer used in the present studies;† there is little likelihood that a g_{VI} function of the same breadth would be applicable to all instruments and to all states of optical alignment.

The reader's attention is directed to Fig. 2 for an illustration of the convolution process. Here the functions considered are $g_{II}=|\epsilon|^{-\frac{1}{2}}$ and $g_{IV}=\exp k_{IV}\epsilon$, where the angular limits are, respectively, 0 to $\epsilon_m=-(\gamma^2/114.6)\cot\theta$ degrees and 0 to $-\infty$. The absorption constant k_{IV} is given by $4\mu l/114.6\sin 2\theta$. Substitution of the representative experimental values $2\theta=24°$, $\gamma=1°$, $l=17$ cm, and $\mu=34$ (moderately absorbing specimen) gives ϵ_m and k_{IV} the approximate values $-0.05°$ and 50, respectively. Application of Eq. (1) to these two functions gives, for the convolution to be evaluated,

$$g_{II,IV}(\epsilon)=\int|\zeta|^{-\frac{1}{2}}[\exp k_{IV}(\epsilon-\zeta)]d\zeta, \qquad (3)$$

which for numerical computation can be replaced by the corresponding summation,

$$g_{II,IV}(\epsilon)=\sum|\zeta|^{-\frac{1}{2}}[\exp k_{IV}(\epsilon-\zeta)]\Delta\zeta. \qquad (4)$$

The accuracy of the numerical results can be made as good as desired by keeping the interval $\Delta\zeta$ sufficiently small. The carrying out of these calculations can be facilitated by tabulating the necessary numerical values of $|\zeta|^{-\frac{1}{2}}$ and $\exp k_{IV}\zeta$ in vertical columns on two strips, which are translated stepwise with respect to each other to permit evaluation of the series at successive values of ϵ. The labor is lightened by utilizing a calculating machine capable of accumulating products. The sharp cutoff in the function g_{II} at $\epsilon=-0.05°$ (see Fig. 2) results from the limited horizontal divergence of the x-ray beam.

III. COMPARISON OF THEORETICAL AND OBSERVED LINE PROFILES

Figure 3 compares theoretical and experimental line profiles at low and intermediate Bragg angles. The experimental profiles were obtained by manually counting over the $(10\cdot1)\beta$ and $(13\cdot2)\alpha_1$ reflections of quartz powder. With Cu*K* radiation these lines occur at the respective angles 24.0 and 90.7°. The x-ray spectrometer was carefully aligned before the measurements and adjusted for the best resolution possible without an undue sacrifice of diffracted power. Sufficient counts were totalized at each point to reduce errors arising from counting statistics to a negligible level. The quartz

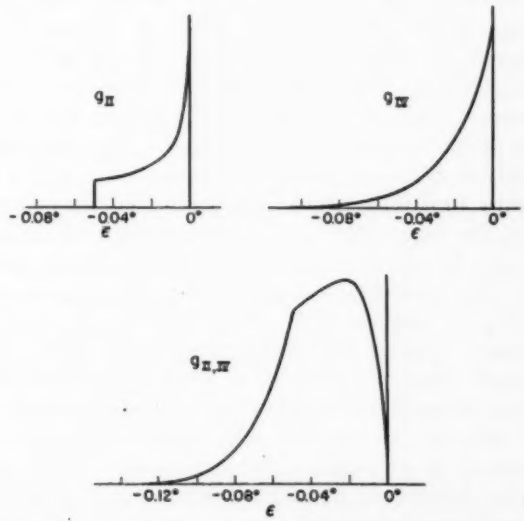


FIG. 2. The instrumental weight functions g_{II} and g_{IV} computed for typical experimental conditions and their convolution, $g_{II,IV}$.

† The instrument employed was a Norelco wide-angle Geiger-counter x-ray spectrometer with narrow source and Soller slits.

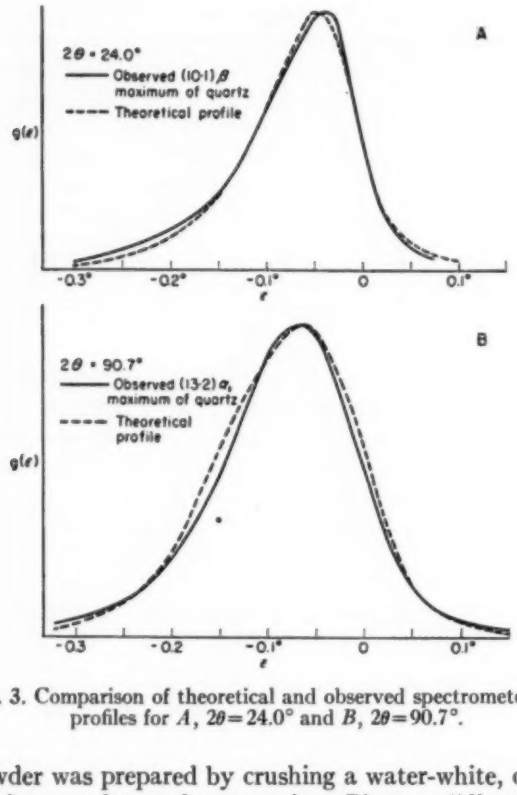


FIG. 3. Comparison of theoretical and observed spectrometer line profiles for A, $2\theta=24.0°$ and B, $2\theta=90.7°$.

powder was prepared by crushing a water-white, clear, single crystal to a fine state in a Plattner "diamond" mortar and separating a size fraction in the range 5–20 microns by sedimentation from methanol. This procedure yields a powder free of particles fine enough to produce line broadening, insuring that the experimental line profiles at large Bragg angles will be the result of purely instrumental factors. Table II gives in detail the experimental conditions that apply to the observed and theoretically computed profiles of Figs. 3A and 3B.

The agreement between theory and experiment displayed by the figures is evidently rather good, especially in view of the fact that the forms of several of the instrumental weight functions cannot be specified with great accuracy. This is particularly true of the vertical-divergence and spectral-impurity functions, the former being very important at 24° and both exerting a significant influence at 90°. Likewise, two of the less weighty functions can be approximated only imperfectly, the source and the sample-absorption functions, the latter because of the difficulty of assaying the precise absorption coefficient μ of the powder, which varies with its state of compactness. The form of the vertical-divergence profile for spectrometers with Soller slits, although already subjected to detailed mathematical analysis, is nevertheless known to but a first approximation because of unusual difficulties which are encountered in attempting to incorporate certain geometrical features into the treatment. These include the finite height of the x-ray source and the "end" effects resulting from the use of a limited number of Soller-slit apertures. It is perhaps not surprising, then, that a misalignment function of 0.04° half-maximum breadth gave improved agreement at 24° but was not needed at 90°. It may be mentioned that at an intermediate angle ($2\theta=50°$) the same misalignment function with $w=0.04°$ again was found to yield improved agreement.

The foregoing discussion of the limitations affecting line-profile studies of a quantitative nature supports the conclusion that further efforts to refine the theoretical treatment would hardly be justifiable. The principal excuse for the very considerable efforts already expended on this topic has been the generally felt need on the part of diffractionists for a clearer understanding of the factors responsible for observed line shapes. Besides contributing to this clearer understanding, it is hoped that these studies will make it possible for workers to apply the x-ray spectrometer to a wide variety of practical problems somewhat more efficiently and productively. One specific practical outcome of the recent studies is treated in the next section.

IV. THE CORRECTION OF X-RAY SPECTROMETER LINE PROFILES FOR INSTRUMENTAL BROADENING

The determination of crystallite sizes from the breadths of x-ray diffraction lines with the aid of the Scherrer equation,[8]

$$D=k\lambda/\beta\cos\theta, \tag{5}$$

is contingent upon one's ability to deduce from an observed line breadth B the pure diffraction broadening β, free of instrumental effects. In a previous publication[9] a brief review was given of current methods for correcting experimental line breadths obtained photographically, and new correction curves were developed by the method of convolutional synthesis for use with spectrometers of the older design without Soller slits,

TABLE II. Experimental conditions governing the line profiles of Fig. 3.

Function		(10·1) $CuK\beta$ $2\theta=24.0°$	(13·2) $CuK\alpha_1$ $2\theta=90.7°$
I. Source	$w=$	0.02°	0.02°
II. Flat sample	$\gamma=$	1°	4°
	$\epsilon_m=$	−0.04°	−0.14°
III. Vertical divergence		Computed for Soller slits and $\delta=2.2°$, $2\theta=24°$	Computed for Soller slits and $\delta=2.2°$, $2\theta=90°$
IV. Absorption	$\mu=$	34	34
	$k_{IV}=$	50	25
V. Receiving slit	$w_s=$	0.025°	0.050°
VI. Misalignment	$w=$	0.04°	Neglected
VII. Spectral impurity	$w=$	Neglected	0.04°

[8] P. Scherrer, Göttinger Nachrichten **2**, 98 (1918).
[9] L. Alexander and H. P. Klug, J. Appl. Phys. **21**, 137 (1950).

having broad sources, and capable of measurements only in the angular range $2\theta=0$ to 90°. Subsequently Schoening, van Niekerk, and Haul,[10] employing the Fourier integral equation, derived a correction curve for spectrometers on the assumption that all experimental line shapes may be described by the function $c/(1+k^2\epsilon^2)^2$. Although they did not allow for geometrical differences between instruments having narrow and wide sources or for the dependence of the geometrical effects upon 2θ, their correction curve agrees closely with that given by Alexander and Klug[9] for spectrometers of the older design.

Figure 4 gives new correction curves, derived in the present convolution studies, which are applicable to spectrometers with narrow sources (rectangular focal spot viewed laterally), Soller slits, and designed for operation over the wider angular range extending from $2\theta=0$ to 165° or more. These curves were derived by a method similar to the one described earlier,[2,9] with the difference that instead of employing successive broadening-ratio curves to arrive at the net breadth resulting from the action of the pure diffraction profile, $f(\epsilon)$, and the several instrumental factors, $g_{\rm I}$, $g_{\rm II}$, $g_{\rm III}$, etc., complete instrumental line profiles, $g(\epsilon)$, were first synthesized for various sets of typical experimental conditions, after which convolutions were executed of these $g(\epsilon)$ profiles with $f(\epsilon)$ profiles of various breadths β (see reference 2, pp. 138–9). As in the earlier work, the basic and rather well justified assumption is made that the pure diffraction profile has the form $1/(1+k^2\epsilon^2)$.

Because of the pronounced differences between the geometrical conditions underlying measurements at small and those at very large Bragg angles, it is necessary to define no less than two correction curves, one for each of these limiting conditions. At low or moderate Bragg angles (2θ between 0 and 90°) the instrumental profile $g(\epsilon)$ is produced by the joint action of a number of factors of approximately equal weight, several of them being asymmetric (see Fig. 1B). The convolutions of such profiles with pure diffraction profiles $f(\epsilon)$ of various breadths lead to the correction curve designated "low-angle reflections." The difference between the positions of the curves for half-maximum and integral breadths is so small as to be justifiably neglected considering other larger sources of inaccuracy that are present. It is felt that this curve is more reliable than those previously published for spectrometers with wide sources[9] because the former dominant role of the x-ray source profile has in the present analysis been superseded by the combined influence of several factors, most of them known with somewhat better accuracy.

At large Bragg angles ($2\theta \geqq 145°$) broadening due to the several instrumental factors becomes small in relation to that resulting from the natural spectral distribution in the incident radiation, as pointed out earlier.

[10] Schoening, van Niekerk, and Haul, Proc. Phys. Soc. (London) **B65**, 528 (1952).

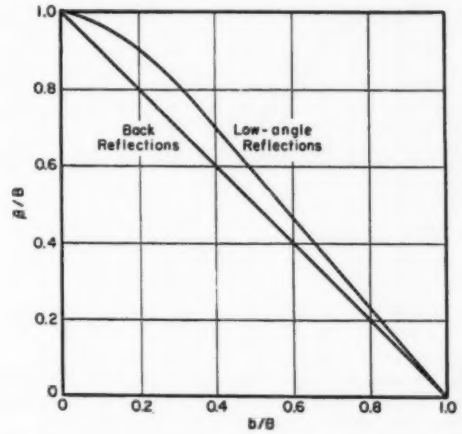


FIG. 4. Curves for correcting x-ray spectrometer line breadths for instrumental broadening. Spectrometers with narrow x-ray sources and Soller slits.

Data compiled by Compton and Allison[6,11] from double-crystal spectrometer measurements indicate that the profiles resulting from the spectral distribution (or impurity) conform rather closely to the function $1/(1+k^2\epsilon^2)$. This means, then, that to a first approximation the observed profile at large Bragg angles is simply the convolution of two $1/(1+k^2\epsilon^2)$ profiles, one due to the crystallite-size distribution and the other to the spectral distribution in the radiation employed. Computation of a series of such convolutions leads to the possibly unexpected result labeled "back reflections" in Fig. 4, a straight line from the upper left-hand to the lower right-hand corner of the chart. This correction applies to both widths at half-maximum intensity and integral breadths. It is equivalent to

$$\beta/B+b/B=1$$

or

$$\beta=B-b, \tag{6}$$

which states that β and b are simply additive, the original Scherrer postulate.[8] It must be strongly emphasized at this point that this correction formula has no general applicability in diffraction measurements because it is strictly valid only when both $f(\epsilon)$ and $g(\epsilon)$ are of the form $1/(1+k^2\epsilon^2)$. For spectrometers with narrow source profiles this condition is approached at back-reflection angles, and it is therefore suggested that the straight-line function be adopted for all crystallite-size measurements in the angular range 145–180°. Actually it is probable that in any particular crystallite-size determination at high angles the optimum correction curve to be applied lies somewhere between the two curves of Fig. 4 but rather closer to the lower one.

The much narrower instrumental profiles attained in the newer spectrometer design permit the determination of β from experimentally measured breadths b and B

[11] A. H. Compton and S. K. Allison, *X-Rays in Theory and Experiment* (D. Van Nostrand Company, Inc., New York, 1935), second edition, Chap. 9.

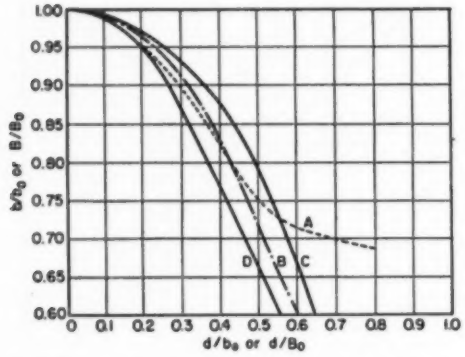


FIG. 5. Curves for correcting line breadths for $K\alpha$-doublet broadening. A is Jones' curve. C and D apply to Gaussian and $1/(1+k^2\epsilon^2)$ line profiles, respectively. B applies to a line of intermediate shape.

with far greater precision than was possible with older spectrometers. This is because the ratio b/B decreases as the instrumental breadth b decreases, thereby causing the point of calculation to move to the left along the curves of Fig. 4, and the precision with which β/B and β can be determined improves toward the left, especially when the calibration curve is nonlinear and, instead, convex upward. It is worthwhile at this point to reaffirm the general rule that for best precision in crystallite-size measurements the experimental conditions should be selected with a view to making b/B as small as possible.

It is also noteworthy that the lower curve of Fig. 4 is of special importance because it applies to the angular region in which $\cos\theta$ has minimum values, thereby permitting measurements of relatively large crystallite dimensions, D [see Eq. (5)]. By employing resolved α_1 or β lines it is possible to do rather accurate work in the size range 500 to 2000A, and size estimates can be made up to about 3000A. Below 500A partial superposition of the α_1 and α_2 profiles sets in, making it difficult to utilize the components of the doublet, and better results will be realized by resorting to a β line or an unresolved $K\alpha$ line in the medium- or low-angle region.

As pointed out in the earlier paper,[9] line broadening resulting from the separation of the $K\alpha$ doublet should be allowed for before the curves of Fig. 4 are applied. Jones[12] showed how such a correction curve can be derived, assuming lines of any known shape. His method consists essentially of compounding two profiles of the same shape and breadth b but differing in height by a factor of two and separated by some angular increment d, and then dividing the resultant area by the resultant height to yield the final *integral* breadth b_0. Assuming a line profile of more or less typical shape, Jones in this way arrived at the correction curve A of Fig. 5, in which b/b_0 is depicted as a function of d/b_0. Of course, the curve applies equally well to broadened diffraction profiles, yielding B/B_0 as a function of d/B_0. In Jones' notation b_0 and B_0 refer to the observed line breadths including the effect of $K\alpha$-doublet broadening, b and B refer to the line breadths after correction for $K\alpha$-doublet broadening, and d is the angular separation of the doublet for the particular radiation employed and at the appropriate angle 2θ.

In order to conform to the usage of the present treatment, which employs breadths at *half-maximum* intensity, the curves B, C, and D of Fig. 5 have been developed by a method similar to that of Jones. Curves C and D are applicable to Gaussian and $1/(1+k^2\epsilon^2)$ profiles, respectively, whereas B has been derived for a profile of intermediate shape. It is suggested that B or D be used for spectrometer line profiles, the latter being more suitable at back-reflection angles. The experimental conditions should be chosen so as to make d/b_0 or d/B_0 as small as possible, preferably less than 0.4. The correct choice of curve to be used becomes more critical and the reliability of the correction poorer as the ratio increases. To measure crystallite sizes with the highest absolute accuracy, it is best to circumvent the α-doublet correction whenever possible by utilizing resolved lines such as α_1, α_2, or β, or, failing in this, a careful analysis of the line shape should be carried out and a more appropriate correction curve derived than those of Fig. 5. Values of the angular separation of the $K\alpha$ doublet as a function of θ for the common $K\alpha$ radiations have been published.[13]

V. ILLUSTRATIVE CRYSTALLITE-SIZE DETERMINATION IN THE BACK-REFLECTION REGION

The specimen was micronized quartzite powder sedimented from methanol to remove all particles larger than 5 microns in mean dimension. The diffraction pattern obtained with Cu$K\alpha$ radiation showed appreciable broadening only in the back-reflection region, so this served as a good test of the new spectrometer with narrow source when applied to the determination of a mean crystallite dimension near the upper limit of the measurable range. The $(24\cdot0)\alpha_1$ line at $2\theta=146.6°$ was counted point by point and the profile compared with a reference profile of the $(24\cdot0)\alpha_1$ line of a 5 to 20-micron fraction of pure quartz powder, the respective widths at

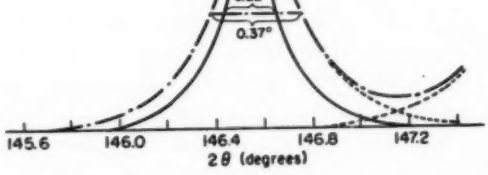


FIG. 6. Shapes of the $(24\cdot0)$ line of quartz powder as observed for the size fractions <5 microns and 5–20 microns with a spectrometer having a narrow x-ray source and Soller slits.

[12] F. W. Jones, Proc. Roy. Soc. (London) **A166**, 16 (1938).

[13] *International Tables for X-ray Crystallography* (Kynoch Press, Birmingham, 1952), Vol. III (in preparation); or *Internationale Tabellen zur Bestimmung von Kristallstrukturen* (Gebrüder Borntraeger, Berlin, 1935), Vol. II, pp. 588–608.

half-maximum intensity being 0.370 and 0.220° (see Fig. 6). The latter line is conspicuously broadened by the natural spectral breadth of the Cu$K\alpha$ radiation, as can be seen by comparing its breadth, 0.220°, with the breadths observed at relatively low angles, which are of the order of 0.10° (see Fig. 3A).

The following numerical results are obtained if the shape factor K of Eq. (5) is assigned a value of 0.90:

$$D_{24.0}=0.9\lambda/\beta\cos\theta$$

$$=\frac{0.9\times1.54\times57.3}{\beta\times0.287}=276/\beta.$$

$b=0.220\pm0.015°$.

$B=0.370\pm0.015°$.

$b/B=0.595\pm0.050$.

$\beta/B=0.405\pm0.050$ (using the back-reflection curve of Fig. 4).

$\beta=0.150°$.

$D_{24.0}=1840$A.

Approximate precision limits in $D_{24.0}$, 1700–2100A.

JOURNAL OF APPLIED PHYSICS VOLUME 25, NUMBER 2 FEBRUARY, 1954

Rocket Measurements of Upper Atmosphere Ambient Temperature and Pressure in the 30- to 75-Kilometer Region*

H. S. Sicinski, N. W. Spencer, and W. G. Dow†
University of Michigan, Engineering Research Institute, Ann Arbor, Michigan

(Received May 19, 1953)

A method for determining ambient temperature and ambient pressure in the upper atmosphere is described, using the properties of a supersonic flow field surrounding a right circular cone. The underlying fundamentals stem from basic aerodynamic principles as combined with the developments of the aerodynamics of supersonic cones by G. I. Taylor, J. W. Maccoll, and A. H. Stone. The experiment provides the necessary cone pressures, velocities and Eulerian angles, such that a Mach number characterizing the ambient space conditions may be computed. A description is given of the requisite experimental equipment and related techniques. Experimental data from two rocket-borne equipments are presented with the resulting calculated pressures and temperatures as experienced over New Mexico to approximately 70 kilometers.

I. INTRODUCTION

THE University of Michigan Department of Electrical Engineering has been engaged for the past several years in the measurement of the ambient temperature and pressure of the upper atmosphere. These measurements have been carried out in high altitude rockets, in particular, the "V-2" and the "Aerobee."

During the period in which V-2 rockets were employed, temperature measurements were implemented by application of the "barometric equation" to a measured curve of ambient pressure *versus* altitude. Although curves of ambient temperature *versus* altitude were obtained,[1] the computational procedure for obtaining the temperature is primarily one of differentiation, and hence yields only very approximate values of temperature.

In an effort to improve the quality of the measurements, a more exact method has been developed which overcomes certain disadvantages of the earlier procedure. Essentially point-by-point values are obtained in a manner that does not require an averaging process. That is, each temperature point on the curve is determined directly from the experimental pressure data independently of other points and gives the temperature at a particular location in space, as contrasted with values obtained from the barometric equation which represent an average over a rather considerable altitude interval.

The experimental data required for a temperature computation by the new method include in general: a ratio of the nose-cone tip (impact) pressure to the pressure at some point on the cone wall, a determination of the instantaneous angle between the rocket's longitudinal axis and a space-fixed reference system, and the magnitude of the missile velocity vector in the same reference system.[2]

The required pressure measurements are accomplished in the missile through the use of "Alphatron" ionization gauges, which are utilized in equipment that

* The research reported in this paper has been sponsored by the Geophysical Research Directorate of Air Force Cambridge Research Center, Air Research and Development Command, under Contract Nos. AF19(122)-55 and W33-038 ac 14050.

† Department of Electrical Engineering, University of Michigan, Ann Arbor, Michigan.

[1] Rpt. No. 2, Upper Air Research Program, Engineering Research Institute, University of Michigan, July 1948.

[2] Another temperature measurement method, similar in that correspondingly fundamental pressure measurements are employed, has been utilized by the Naval Research Laboratory. See Haven, Koll, and La Gow, J. Geophys. Research 57, 59–72 (1952).

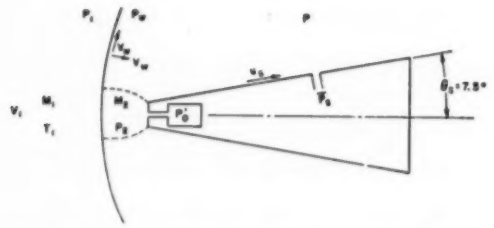


FIG. 1. Physical quantities appearing in non-yaw conical flow. Pressures measured by the experiments described appear inside of the outline of the cone. (For physical significance of symbols, see list in paper.)

has been developed by this research group. The data obtained is telemetered from the rocket to ground stations where it is recorded.

The fundamental information required for angle computation is similarly determined in the missile through the use of a single gyroscope. In this case, the data are recorded in the rocket on film, which is later recovered when the missile reaches the ground.

Velocity information is obtained by triangulation employing ground based instrumentation which tracks the rocket during flight.

Temperature measurements have been made on several rocket flights utilizing the new method. The following sections of this paper present the data resulting from two such flights of Aerobee rockets, a discussion of the theoretical basis for the measurements and a description of the particular equipment developed to obtain the basic data.

II. THEORETICAL CONSIDERATIONS

A. General

Temperature is a typical "intensive" magnitude[3] quantity, which, for its determination, must be correlated with phenomena measured "extensively." Although the usual laboratory thermometric systems can measure temperature in extensive terms, these techniques cannot be directly extrapolated to supersonic missiles for upper atmosphere ambient temperature measurements without producing questionable results. The chief difficulty arises with the formation of a boundary layer about the instrument, which perturbs the temperature experienced. A more promising datum is pressure, which, unlike the temperature, is very nearly constant throughout any boundary layer section, being nearly equal to the value just outside the boundary layer. The pressure datum thus "neglects" the boundary layer, approximating an inviscid flow.

For the practical case of a cone with a semi-vertex angle of 7.5°, very good agreement exists[4] between the inviscid theory and experimental data from viscid tests in the range of Mach numbers and yaw angles experienced.

The principal limitation to applying inviscid theory occurs for large yaw angles ($\epsilon \sim 0.8\theta_s$) where boundary layer separation occurs.[5]

B. Non-Yawing Cone

The problem of supersonic flow around a cone has been successfully analyzed by Taylor and Maccoll,[6] with the subsequent embodiment of their results in tabular form by Z. Kopal.[7] These results are applicable to cone-pressure measurements in order to compute ambient upper air conditions for given conical geometry and the characteristic Mach numbers. Although surface pressure measurements alone do not constitute sufficient information to deduce the characteristic Mach number, knowledge of the total head pressure will, when taken concurrently with the surface pressures, define the characteristic Mach number. A schematic representation of the experiment and the physical quantities appearing is shown in Fig. 1. Quantities with subscript "one" are the upstream conditions or so-called "ambient values." The dotted area about the cone vertex defines a subsonic flow region existing because the cone is truncated to permit total head measurement. This deviation from purely conical geometry affects the flow locally; however, a wind-tunnel analysis demonstrates that the conical flow regime is established ahead of the cone surface pressure measuring port.

The computation of the ambient conditions proceeds through a combination of the Taylor-Maccoll relations and the Rayleigh total-head expression. P_0' and $\bar{P}_s$ (cone-tip and cone-wall pressures) and cone velocity V relative to the ambient air are measured experimentally. The theory presented leads to a relationship (Fig. 2)

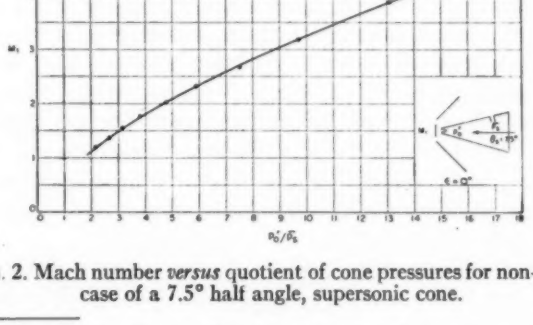


FIG. 2. Mach number *versus* quotient of cone pressures for non-yaw case of a 7.5° half angle, supersonic cone.

[3] For definition of extensive and intensive properties see F. E. Fowle, *Smithsonian Physical Tables*.

[4] Cronvich and Bird, *Pressure Distribution Tests for Basic Conical Flow Research* (Ordnance Aerophysics Laboratory, Daingerfield, Texas).

[5] Franklin K. Moore, "Laminar boundary layer on a cone in supersonic flow at large angles of attack," NACA-TN-2844.

[6] G. I. Taylor and J. W. Maccoll, Proc. Roy. Soc. (London) A139, 279–311 (1933). J. W. Maccoll, Proc. Roy. Soc. (London) A159, 459–472 (1937).

[7] Z. Kopal, Massachusetts Institute of Technology Tech. Report No. 1, 1947, Department of Electrical Engineering, Center of Analysis.

between $P_0'/\bar{P}_s$ and the Mach number M_1, thus permitting a determination of the Mach number from the experimental data. The dependence of Mach number and cone-wall pressure $\bar{P}_s$ on the ambient pressure P_1 appears in the course of this determination. The ambient temperature is determinable from the familiar Eq. (6) relationship between Mach number and velocity relative to ambient air.

It is convenient to initiate the theoretical analysis by stating that ratio $P_0'/\bar{P}_s$ of the measured pressures in the following identity:‡

$$\frac{P_0'}{\bar{P}_s}=\frac{P_0'}{P_1}\frac{P_1}{P_w}\frac{P_w}{P_0}\frac{P_0}{\bar{P}_s}. \tag{1}$$

The theoretical treatment consists in expressing each of the right-hand factors in terms of the Mach number and the known ratio of specific heats, thus leading to the Fig. 2 relationship.

To accomplish this for the first factor, energy considerations permit expressing the ratio of pressures across the normal shock wave in terms of the Mach number and the ratio of specific heats as follows:‡

$$\frac{P_1}{P_0'}=\left[\frac{2\gamma M_1^2-(\gamma-1)}{(\gamma+1)}\right]^{(1/(\gamma-1))}\left[\frac{(\gamma+1)M_1^2}{2}\right]^{\gamma/(1-\sigma)}. \tag{2}$$

Similar evaluation of the remaining three factors on the right of Eq. (1) requires use of the theory of the conical regime. Taylor and Maccoll[8] determined the pressure ratio across the shock wave using a lengthy graphical procedure. Kopal[7] derived an explicit expression for this ratio using purely algebraic procedures; he also prepared tables providing values of the local velocities (radial velocity U_w, tangential velocity V_w, and sonic velocity, a) for any given cone angle as a function of the Mach number. By using values from these tables in his expression,

$$\frac{P_w}{P_1}=\frac{(\gamma^2-1)(c^2-u_w^2-v_w^2)}{4\gamma v_w^2-(\gamma-1)^2(c^2-u_w^2-v_w^2)}, \tag{3}$$

the second factor on the right of Eq. (1) can be evaluated in terms of Mach number and γ.

The quantity c, appearing here, is a useful reference velocity, sometimes defined as the maximum velocity attainable by converting all the heat energy of the fluid into uniform motion. In terms of the local sonic velocity, a, and the velocity V relative to ambient air, the reference velocity c is defined as

$$c^2=V^2\left[1+\frac{2a^2}{V^2(\gamma-1)}\right]. \tag{4}$$

‡ See list of symbols at end of paper.

[8] G. I. Taylor and J. W. Maccoll, Proc. Roy. Soc. (London) A139, 288–292 (1933).

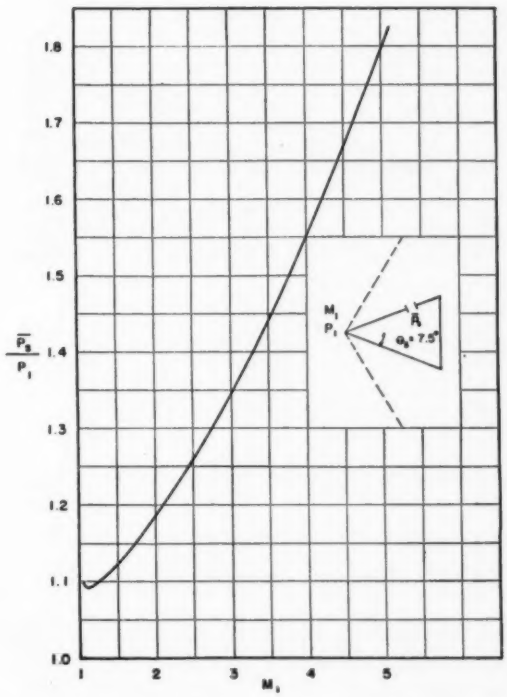


FIG. 3. The relationship between the unyawed surface pressure and the ambient pressure as a function of the free stream Mach number for a non-yawing, 7.5° half angle supersonic cone.

The third factor on the right-hand side of Eq. (1), the ratio of static pressure behind the shock wave to the stagnation pressure, is found directly from the Bernoulli integral and the assumption of adiabatic flow behind the shock surface.[7] This ratio, in terms of the local velocities provided by Kopal's tables, is

$$P_w/P_0=[1-(u_w/c)^2-(v_w/c)^2]^{\gamma/(\gamma-1)}. \tag{5}$$

Lastly, the ratio $\bar{P}_s/P_0$ of cone surface pressure to stagnation pressure behind the shock wave follows from Eq. (5) on setting the tangential velocity (v_w) equal to zero.

The results from using these four evaluation procedures in Eq. (1), expressed in terms of the Mach number for the non-yaw case of a 7.5° half-angle cone, are shown in Fig. 2.

With the requisite Mach number known from Fig. 2, the ambient pressure is available from Eq. (1) after dividing both sides by P_0'/P_1. The result of this procedure is shown in Fig. 3.

Computation of the ambient temperature depends on the definition of the Mach number and the adiabatic sonic velocity relationship. These express the ambient temperature explicitly as

$$T_1=(V/M_1)^2(R\gamma)^{-1}. \tag{6}$$

The velocity V appearing here is the relative flight speed. In the experiments presented, no provision was made for estimating local wind conditions; consequently the missile velocity relative to the earth was taken as

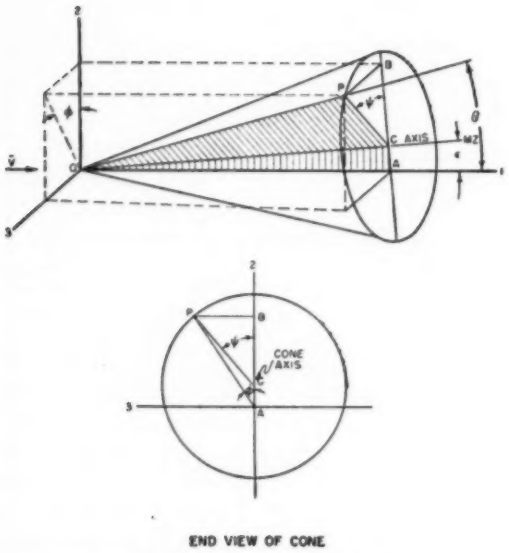


FIG. 4. Coordinates describing the yawing cone.

the defining velocity. Tacitly, this statement assumes that the winds present are negligible compared to the relative flight speed.

These data have given the temperatures through which a reasonably smooth curve could be drawn (Fig. 11). The departure of the measured temperature from the smooth curve are of the nature that would result from the presence of a wind field varying in speed and direction. Uniform wind fields, on the other hand, would yield temperatures, continuously higher or lower than the actual ambient temperature.

C. Yawing Cone

Under the conditions experienced by high-speed vehicles missiles are fundamentally yawing bodies; that is, the missiles longitudinal axis does not usually maintain coincidence with the free stream relative velocity vector. Consequently, the experiment must employ a yawing-cone theory for its analysis.

A theory for yawing supersonic cones was developed by A. H. Stone[9] which included second order yaw effects. Stone's analysis led to Kopal's[10] tabulation of the perturbation coefficients for use in the solution of supersonic flow fields about large-yaw cones. Of particular interest is Stone's expression for the cone's surface pressure, since it provides the basis for a pressure experiment on a yawing cone. In terms of the coordinates of Fig. 4, the surface pressure P_σ is

$$P_\sigma=\bar{P}_s+\epsilon\sum_{n=0}^{\infty}\eta_n\cos n\phi+\epsilon^2\sum_{n=0}^{\infty}P_n\cos n\phi\cdots. \quad (7)$$

$\bar{P}_s$ is the surface pressure for a zero-yaw angle ($\epsilon=0$), while the perturbation coefficients η_n and P_n are available from reference 7. Stone's analysis demonstrated that all the second-order yaw terms except $n=0$ and $n=2$ vanish under the boundary conditions on the cone, while the Rankine-Hugoniot conditions reduce all the first order terms to zero except $n=1$.

In application of Eq. (7) the perturbation coefficients as given by Kopal need to undergo a transformation by means of a Taylor expansion for utilization in the desired reference frame.[11] Kopal's tabulation presents these coefficients relative to arguments of the unyawed reference system where the desired coefficients are perturbed by angle $\langle\Delta\bar{\theta}\rangle$. Thus the pressure at any perturbed position θ is given by the relation

$$P(\theta)=P(\bar{\theta}+\langle\Delta\bar{\theta}\rangle)=P(\bar{\theta})+P'(\bar{\theta})\langle\Delta\bar{\theta}\rangle+P''(\bar{\theta})\langle\Delta\bar{\theta}\rangle/2\cdots, \quad (8)$$

where the barred quantities are with respect to the unyawed reference frame. The primes refer to differentiation with respect to $\bar{\theta}$. With Eq. (8) we still need an expression for the yawed conical surface with respect to the unyawed reference frame. In the region between the shock wave surface and the yawing cone the variables will be constant over surfaces of a generally conical nature. If it is assumed that these surfaces remain cones of circular or elliptical section, rotated through the yaw angle ϵ, the equation[12] for any of these surfaces is

$$\bar{\theta}_s+\Delta\bar{\theta}_s=\bar{\theta}_s+\epsilon\cos\phi-(1/2)\epsilon^2\cot\bar{\theta}_s\sin^2\phi\cdots, \quad (9)$$

where the particular surface is defined when $\bar{\theta}_s$ and $\Delta\bar{\theta}_s$ are stated. Using Eqs. (8) and (9) in Eq. (7), then collecting terms to the order ϵ^2, and evaluating the derivatives gives the pressure at the yawed solid cone surface as,

$$P_\sigma=\bar{P}_s\left[1+\epsilon\cos\phi\left(\frac{\eta}{\bar{P}_s}\right)+\epsilon^2\cos 2\phi\left(\frac{P_2}{\bar{P}_s}+\frac{\gamma}{2}\left(\frac{u_s}{a}\right)^2\right)+\epsilon^2\left(\frac{P_0}{\bar{P}_s}+\frac{\gamma}{2}\left(\frac{u_s}{a}\right)^2\right)\cdots\right]. \quad (10)$$

The experimental data provide values for P_σ, ϵ, and ϕ. These quantities along with Eq. (10) permit computation of $\bar{P}_s$, the unyawed pressure. Having $\bar{P}_s$, the experiment reduces to the non-yaw case for which expressions involving the ambient pressure and temperature have already been given. P_σ are surface pressures measured by suitable gages located in the cone, and ϵ and ϕ are computed from a combination of the trajectory locus and data from a missile-borne gyroscope. From the definitions, the yaw-angle computation makes prerequisite an assumption regarding environmental winds. In both experiments presented the wind velocities are

[9] A. H. Stone, J. Math. Phys. **30**, 200 (1952).
[10] Z. Kopal, Report No. 3, Massachusetts Institute of Technology Department of Electrical Engineering Center of Analysis and Z. Kopal, Report No. 5, 1949, Massachusetts Institute of Technology Dept. of Electrical Engineering Center of Analysis.
[11] Van Dyke, Young, and Siska, J. Aeronaut. Sci. **18**, 355 (1951).
[12] A. H. Stone, J. Math. Phys. **27**, 73 (1948), Eq. 34.

assumed to be small compared to the missile velocity. The yaw angle is then determined with the wind vector tangent to the missile trajectory while the missile aspect is taken from a gyroscope.

III. INSTRUMENTATION

The instrumentation that has been developed and used in Aerobee rockets by this research group to obtain the fundamental data required utilizes an Alphatron gauge as the basic pressure sensitive element, and a gyroscope for missile angular position determination.

This section of the paper will describe briefly the manner in which each of these devices is employed.

A. Pressure Measurement

An Alphatron is an ionization gauge wherein the ionizing energy is obtained from alpha particles emanating from a small quantity of radium, generally of the order of milligrams.

The essential external characteristics of the particular gage chosen for use in this investigation are illustrated in Fig. 5, which presents a typical curve of output current *versus* chamber pressure. The lowest measurable pressure is determined fundamentally by the "dark current," the value on the curve to which the lower portion of the curve is asymptotic. The upper limit is determined by recombination of ionized particles before the ionization products are collected and measured, as evidenced by the bending of the curve at the higher currents.

Two features important from the standpoint of circuit requirements are immediately apparent from the curve: the very small current that constitutes the basic information signal, and the rather large ranges of current and pressure which must be accommodated in an instrument which utilizes the device over its useful range. The small current implies either the use of a relatively small (few megohms) load resistance and very large voltage amplifications, or the use of very high values of resistance, with the consequent problem of impedance matching. The extensive useful range on the other hand demands the use of several subranges in order to obtain a reasonable definition in the ultimate pressure data.

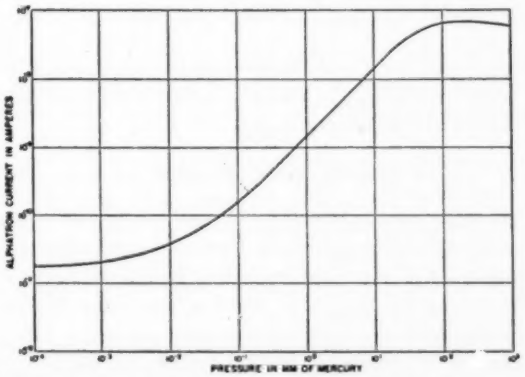


FIG. 5. Variation of output current with chamber pressure for a particular Alphatron pressure gauge.

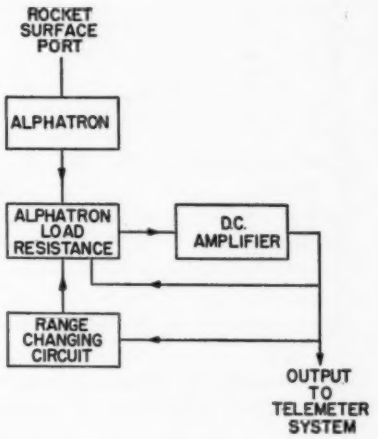


FIG. 6. Elementary block diagram of Alphatron pressure measurement system.

Figure 6 is an elementary block diagram illustrating the circuit developed[13] to meet these requirements. The Alphatron current is passed through a resistance of sufficient magnitude to produce a voltage equivalent to the desired information signal. Because of the very small current, this resistance may be as high as 250 000 megohms.

The voltage obtained across this resistance is applied to a 100 percent negative feedback dc amplifier which acts essentially as an impedance changing device. The first stage of the amplifier employs an electrometer tube in order to provide an input resistance that is large compared with any probable Alphatron load resistance. The following sections of the amplifier are, in sequence: a voltage amplifier, a heater voltage regulator, another voltage amplifier and finally a cathode follower stage. Inside the feedback loop the voltage gain is high, of the order of 4500. However, with feedback, the voltage gain of the system is unity, whereas the current gain is significantly high. Since the output voltage is equivalent to the input voltage (100 percent feedback) the current gain is numerically equivalent to the ratio of the input load resistor (Alphatron load) to the cathode resistor of the cathode follower.

The voltage obtained from the cathode follower constitutes the desired data and is accordingly applied to the recording system, in this case the telemetering system.

In order to provide for the several subranges, different possible values of Alphatron load resistance are provided, one for each subrange. It is the function of the range changing circuit to select and insert the particular load resistance appropriate to a particular range of chamber pressure. To accomplish this, the amplifier output signal is applied to the range changing circuit which uses two thyratrons to control a bi-directional rotary solenoid. If the information signal voltage exceeds a predetermined value (in either direction) the

[13] Developed from an original design by J. R. Downing and G. Mellen, Rev. Sci. Instr. **17**, 218 (1946).

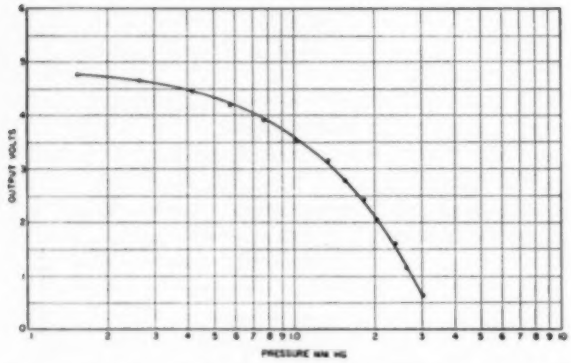


FIG. 7. Variation of Alphatron pressure measurement system output with pressure for a particular sub-range.

next lower or higher value of resistance is inserted, thus returning the information signal to an "on scale" value. An automatic range selection device is of course necessary because the equipment operates unattended through the total pressure range encountered during a rocket flight.

Figure 7 illustrates the variation of output signal with pressure for a particular subrange. In this case the Alphatron load resistance is 5000 megohms.

The complete Alphatron equipment in a particular rocket includes several nearly identical, independent units similar to that described above, differing only perhaps in regard to choice of pressure subranges as may be required by particular gauge locations, for example, cone-wall or cone-vertex mounting.

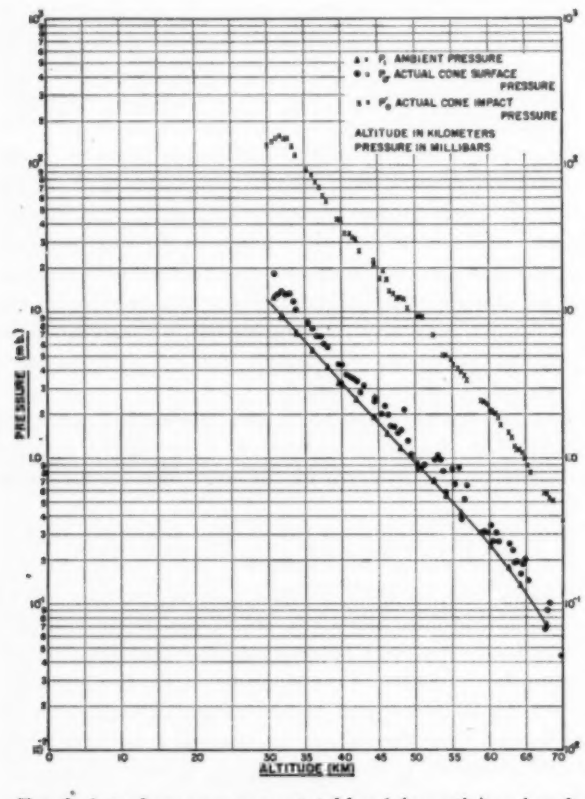

▲ • P_i AMBIENT PRESSURE
● • P_p ACTUAL CONE SURFACE PRESSURE
x • P'_0 ACTUAL CONE IMPACT PRESSURE
ALTITUDE IN KILOMETERS
PRESSURE IN MILLIBARS
PRESSURE (mb.)
ALTITUDE (KM)

FIG. 8. Actual cone pressures, total head (cone tip) and surface, compared with the ambient pressure for Aerobee rocket of June 20, 1950 at 0838 hours, at Holloman Air Force Base, Alamogordo, New Mexico.

B. Angle Measurement

The gyroscope used for missile angular position measurement is a modified Sperry type F4A unit. The modification was accomplished primarily in order to allow operation under free fall conditions. However, in addition, an attachment was developed that enables the establishment of zero position prior to rocket flight.

Recording of gyroscope data is accomplished by photography of the gyroscope sphere (gyrostat) posi-

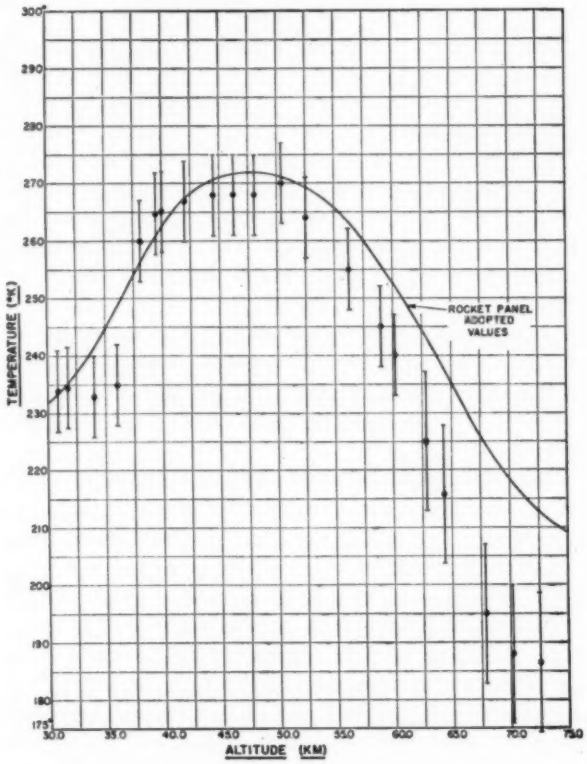


FIG. 9. Ambient temperature at various altitudes above Alamogordo, New Mexico. These temperatures are computed from Aerobee rocket data of June 20, 1950 at 0838 hours using the assumption of a limited wind field.

tion in reference to a missile-fixed coordinate system. The film on which the position is recorded is recovered at the end of the rocket flight.

IV. EXPERIMENTAL RESULTS

Two experiments based on the above theoretics have been successfully completed. Pressure equipments were instrumented in Aerobee Sounding Rocket type missiles for launching by the U. S. Air Forces at the Holloman Air Force Base at Alamogordo, New Mexico. The first experiment on June 20, 1950, carried one impact pressure gauge and one cone-surface pressure gauge. For comparison purposes the two experimental pressures and

the resulting computed ambient pressure are shown in Fig. 8. From this first experiment, the ambient pressure data is reliable to one part in thirty-five.

The seemingly relatively large scatter in the cone surface pressure is a result of the missile's rotation as it assumes increasing yaw angles with altitude. The impact pressure is generally without such cyclic variations since it remains independent of the yaw angle for values up to about thirty degrees. Although not shown in entirety, maxima in the ratios of both cone surface and impact pressure to ambient pressure occur in the neighborhood of 35 kilometers altitude. These maxima are presumably the result of the combination of maxima in the Mach number and the missile velocity; as such they can only be construed as caused by missile behavior, not representing properties of the atmosphere.

The temperatures computed from these data are shown in Fig. 9. No comment is offered on this curve other than that the maximum probable error is believed to be ±eight degrees Kelvin to about sixty kilometers and ±thirteen degrees Kelvin above about sixty kilometers.

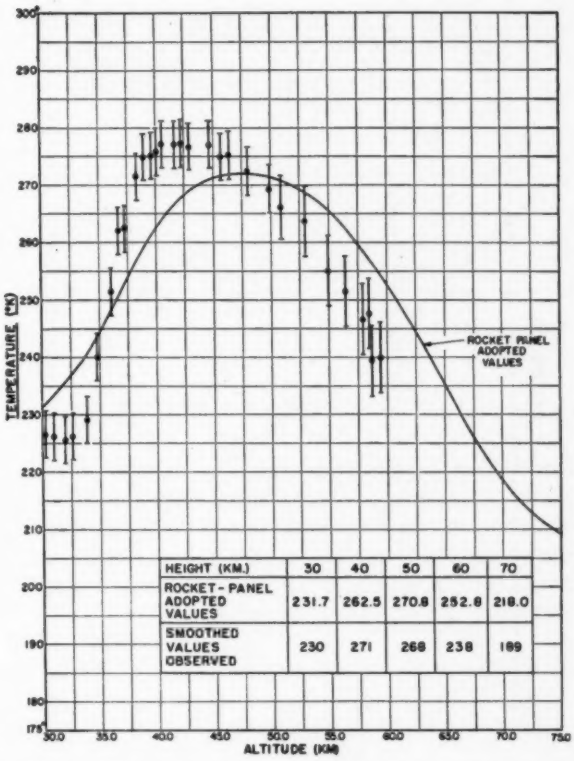
The second experiment was completed on September 13, 1951. The instrumentation represented considerable improvement over that of June 20, 1950, in having two cone-surface gauges and a greatly increased information reporting capacity. The increase in sampling information rate resulted in a reduction of the overall probable error of the final temperature data. For Fig. 10 the temperatures up to fifty kilometers have a maximum probable error of ±five degrees Kelvin while the probable error above fifty kilometers is ±seven degrees. The ambient pressure resulting from this flight has an improved accuracy such that it is reliable to one part in sixty-five. These data are shown in Fig. 11 as the "experimental points."

V. SELF-CONSISTENCY IN THE RESULTS

It has been pointed out above that each experimental point, for the temperature curves of Figs. 12 and 13, is evaluated from the experimental data independently of other points. Furthermore, in the non-yaw case, the temperature calculations leading to Figs. 11 and 12 employ only the ratio of the measured total-head-pressure and cone-surface pressures, not their absolute values. Thus, the absolute values of the pressure measurements are not employed in determining temperatures.

It is obviously possible to employ the experimentally-determined temperatures in the familiar hydrostatic equation[14] thereby determining the absolute values of the ambient pressures by a method that does not directly use the pressure measurements obtained by the rocket instrumentation. In such use of the hydrostatic equation an inverse square variation of gravity is as-

[14] S. K. Mitra, "The Upper Atmosphere," *The Royal Society of Bengal Monograph Series* (1947), Vol. V, Eq. 3, p. 5 ($dP/P = -mgdh/kT_1$).



HEIGHT (KM.)	30	40	50	60	70
ROCKET-PANEL ADOPTED VALUES	231.7	262.5	270.8	252.8	218.0
SMOOTHED VALUES OBSERVED	230	271	268	238	189

FIG. 10. Ambient temperature at various altitudes above Alamogordo, New Mexico. These temperatures are computed from Aerobee rocket data of September 13, 1951 at 0437 hours using the assumption of a limited wind field.

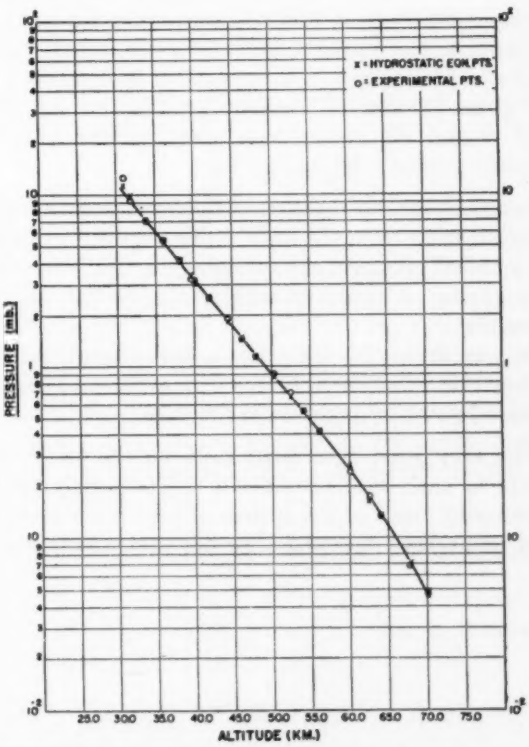


FIG. 11. Pressures calculated from hydrostatic equation ($dP/dh = -Pgm/RT$) using experimental temperature of Fig. 9, compared with ambient pressures for the same flight.

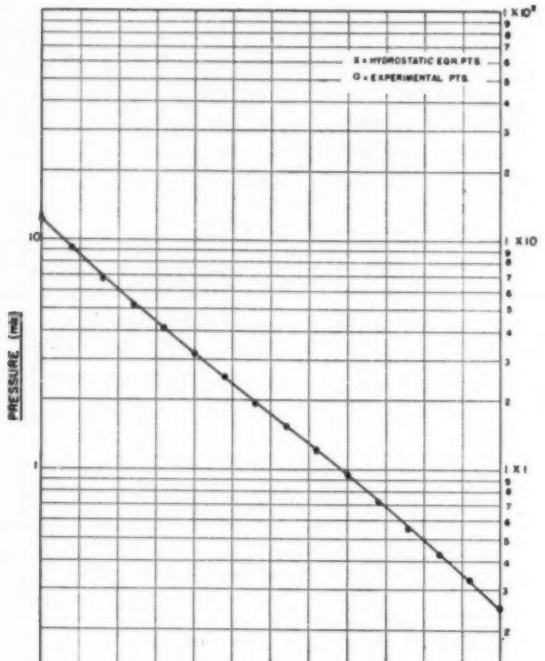


FIG. 12. Pressures calculated from hydrostatic equation using experimental temperatures of Fig. 10, compared with ambient pressures for the same flight.

sumed, and a mean molecular weight of 28.966 for air is employed. Balloon observation results are employed to provide the absolute value of pressure at 30 kilometers, which serves as a constant of integration.

Figures 11 and 12 present, for the two data sets, Figs. 9 and 10, respectively, the absolute values of ambient pressure determined by

(a) As shown by the circles, by direct point-by-point determination from the data, using Fig. 3, in this case the ambient pressures are obtained practically speaking, by applying an appropriate correction to the cone-wall pressures.

(b) As shown by the crosses, by employing in the hydrostatic equation the Figs. 9 and 10 temperatures obtained point-by-point from the data.

The very good agreement between the two sets of points in each case provides a rather satisfying self-consistency check of the system of instrumentation and data reduction employed. The deviations between the two sets of points are less than the experimental errors of the method. Of course, self-consistency as between these two different rocket flights and observations by other methods[15] is also of interest. However, seasonal and diurnal variations of the temperature curve certainly exist and must be taken into account in comparing the two sets of results here reported with one another and with results obtained by other methods.

We wish to take this opportunity to thank Mr. Ralph E. Phinney for his interest and valuable discussions involving the basic aerodynamics of this problem.

LIST OF SYMBOLS

P_ϵ	Cone surface pressure of yawed cone, along ray at angle ϕ from plane of yaw
P_1	Ambient pressure
P_0	Stagnation pressure behind conical shock wave
P_0'	Stagnation pressure behind normal shock wave
P_2	Static pressure behind normal shock wave
P_w	Static pressure at shock wave surface on downstream side of a conical shock wave
$\bar{P}_s$	Cone surface pressure when yaw angle is zero ($\epsilon=0$)
T_1	Ambient temperature
M_1	Free stream Mach number
M_2	Mach number behind normal shock wave
V	Free stream velocity
v_w	Tangential particle velocity
u_w	Radial particle velocity
u_s	Cone surface particle velocity
θ_s	Cone half angle
c	See Eq. (4)
ϵ	Yaw angle
$\eta/\bar{P}_s$	Stone's first-order perturbation coefficient
$P_1/\bar{P}_s$	Stone's second-order perturbation coefficient
$P_0/\bar{P}_s$	Stone's second-order perturbation coefficient
ϕ	Spherical coordinate
θ	Spherical coordinate
ψ	Angle of rotation about cone's longitudinal axis measured from plane defined by cone's longitudinal axis and the wind velocity vector $\bar{V}$
α	Local sonic velocity
h	Altitude
g	Acceleration of gravity
m	Mean molecular weight
R	Universal gas constant
γ	Ratio of specific heats

[15] The Rocket Panel, Phys. Rev. 88, 1027 (1952).

JOURNAL OF APPLIED PHYSICS VOLUME 25, NUMBER 2 FEBRUARY, 1954

Penetration of an Electromagnetic Wave into a Ferromagnetic Material

A. PAPOULIS*
Magnetics Department, Burroughs Corporation Research Activity, Philadelphia, Pennsylvania
(Received June 12, 1953)

When a magnetic field is applied to the surface of a lamination of a metallic ferromagnetic core, it does not instantly penetrate into the interior because of the induced eddy currents. If the B-H curve of the core material is a straight line, the computation of the field within the material can be performed with known methods. For a general B-H curve the resulting equation for H is nonlinear, and only a numerical solution is possible. The purpose of this paper is to give an approximate solution for the case of a rectangular hysteresis loop material. The theoretical results that are developed are in close agreement with the experimental measurements described in this paper.

I. INTRODUCTION

RECENTLY, rectangular loop ferromagnetic materials having a B-H curve similar to that shown in Fig. 1 have found wide application in the digital computing field as storage devices, shift registers, and logical elements. As used in computer applications, the magnetic core consists of a toroidal tape of thickness d, height h, and radius r. The thickness is on the order of a thousandth of an inch, the height, about an eighth of an inch, and the radius, on the order of an inch. The core is usually driven by a constant current source and the output winding is connected to a load. Because of the nonlinearity of the B-H curve, an exact description of the field inside the core is not possible. In the following an approximate solution will be given for a core under various loading conditions.

II. THE FIELD EQUATIONS

1. Assumptions

It will be assumed that the applied magnetizing field H is the z direction (see Fig. 2) and symmetrical around the circumference of the tape; this assumption is satisfactory even if the driving coil covers only part of the core, since the permeability of the material is much larger than the permeability of the air. Since $d \ll r$, H will be assumed to be the same on both sides of the tape. The induced field E is in the y direction. Thus, E-H will be treated as a plane wave independent of y and varying only in the x direction, as in the case of a magnetic sheet of thickness d that is infinite in y and z, with a uniform field H applied at its surfaces $x=0$ and $x=d$.

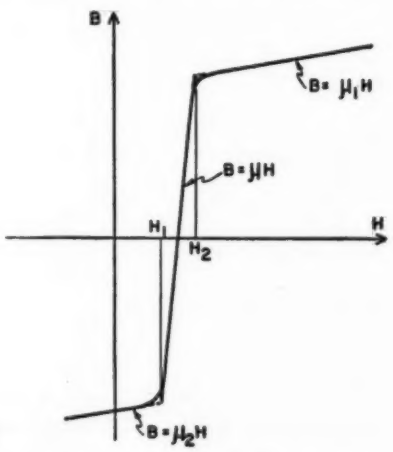


FIG. 1. Typical dc hysteresis curve for increasing H.

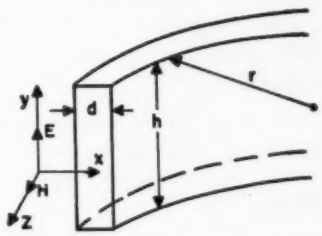


FIG. 2. A section of the metallic ribbon of a magnetic core.

If the dielectric constant ϵ is neglected Maxwell's equations take the form

$$\partial E/\partial x = -\partial B/\partial t, \tag{1}$$

$$\partial H/\partial x = -\sigma E, \tag{2}$$

hence

$$\partial^2 H/\partial x^2 = -\sigma \partial B/\partial t. \tag{3}$$

In the Eqs. (1)–(3) E is a scalar in the y direction, H a scalar in the z direction, σ is the conductivity of the material, and B is a known function of H. It is assumed that at every point within the metal, B takes instantly the value $B(H)$ given by the dc hysteresis loop for increasing H, where H is the magnetic field at the point under consideration.

2. The Initial and Boundary Conditions

The magnetizing field H is symmetrical about $d/2$; hence its slope, $\partial H/\partial x = -\sigma E$, which is continuous, must equal zero for every t at $x=d/2$. Thus

$$E(d/2,t)=0. \tag{4}$$

If the driving current $i(t)$ is given, then

$$H(0,t)=H_0(t)=Ni(t)/s, \tag{5}$$

* Polytechnic Institute of Brooklyn, Brooklyn, New York; Consultant, Burroughs Corporation Research Center, Philadelphia, Pennsylvania.

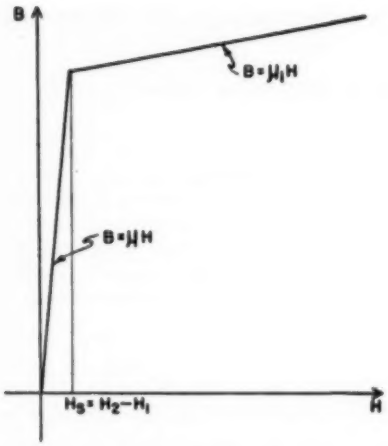


FIG. 3. Idealized dc hysteresis curve for increasing H.

where $s=2\pi r$ is the circumference of the material and N is the number of turns on the driving coil. The core consists of w wraps of the tape; in the following it will be assumed that $w=1$; with obvious modifications the results can be extended to any value of w.

The integral of E around the path of length $2h$ encircling the core equals the rate of change of the flux ϕ therefore

$$E(0,t)=\frac{1}{2h}\frac{d\phi}{dt}=\frac{e(t)}{2hN}, \tag{6}$$

where

$$e(t)=Nd\phi/dt$$

is the induced voltage in the driving coil. In general $i(t)$ and $e(t)$ satisfy a relation depending on the source and on the loading of the core. Thus, at $x=0$ a known relation must hold between H and E. Special cases will be considered below in the discussion of the core under load.

We assume H in the interior of the core to be zero at $t=0$ (at the moment the driving current is applied); thus

$$H(x,0)=0 \quad 0<x<d, \tag{7}$$

3. The Square-Loop Magnetic Material

Figure 1 shows a typical B-H curve for the type of material under consideration. It can be approximated by three straight lines; μ_1 and μ_2 are small compared to μ. In the following, μ_2 will be taken equal to zero; the error due to this assumption is small if the applied H is large compared to H_1. The B-H curve will thus be approximated as shown in Fig. 3. It will be assumed initially that $\mu_1=0$; the effect of μ_1 will be evaluated in a later section.

Plotting H and E as functions of x for a given time t, the curves of Fig. 4 are obtained. The intersection of the H curve with the line $H=H_s$ determines the point x_s. For $0\leqq x\leqq x_s$, H is greater than H_s; hence in this interval the core is saturated and $B=\mu H_s=$ constant. Therefore (see Eq. (1)) $\partial E/\partial x=0$ and $E=E_s(t)$ = constant for a fixed t. Integrating (2) from 0 to x_s, we obtain Eq. (8) since $H(x_s)=H_s$ and E is constant in this interval.

$$H(0,t)-H_s=\sigma x_s(t)E_s(t) \tag{8}$$

hence

$$E_s(t)=E(0,t)=\frac{H(0,t)-H_s}{\sigma x_s(t)}. \tag{9}$$

We shall next determine $dx_s(t)/dt$ in terms of $E_s(t)$. An approximate expression can be derived as follows. Figure 5 shows the variation of B inside one half of the material at a given time t. In the saturated region $0\leq x\leq x_s$, $B=\mu H_s$. The total area $(OAKL)$ under the B curve in the interval $(0,d/2)$ equals $\phi(t)/2h$; similarly, the area of $OAK'L'$ equals $\phi(t+\Delta t)/2h$. If $K'L'$ is considered as a parallel displacement of KL by Δx_s, the area between KL and $K'L'$ (shaded area in Fig. 5) equals $\mu H_s\Delta x_s$; hence in the interval of time Δt, ϕ increases by $2h\mu H_s\Delta x_s$. Therefore (see Eq. (6),

$$E_s(t)=E(0,t)=\frac{1}{2h}\frac{d\phi}{dt}=\frac{\mu H_s dx_s}{dt}$$

or

$$dx_s/dt=E_s(t)/\mu H_s. \tag{10}$$

Equation (10) states that the velocity with which x_s moves inside the metal is proportional to the rate of change of the flux ϕ. Actually the ratio between $E_s(t)/\mu H_s$ and dx_s/dt is a function of time $k(t)$ so that

$$k(t)\frac{dx_s}{dt}=\frac{E_s(t)}{\mu H_s}. \tag{11}$$[1]

If Eq. (11) is integrated from 0 to t_s, the time for which $x_s=d/2$, then it follows that

$$\int_0^{t_s} k(t)\frac{dx_s}{dt}dt=\frac{d}{2},$$

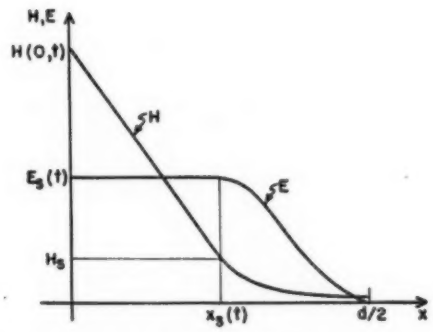


FIG. 4. Variations of E and H inside the ferromagnetic core for a fixed time t.

[1] It can easily be shown that if $H_s=0$ $k(t)=1$; in this case $B=0$ for $x>x_s$.

since

$$\int_0^{t_s}\frac{E_s(t)}{\mu H_s}=\frac{1}{\mu H_s 2h}\int_0^{t_s}\frac{d\phi}{dt}dt=\frac{\phi(t_s)}{\mu H_s 2h}=\frac{d}{2}.$$

Thus the weighted average of $k(t)$ must equal one because

$$\int_0^{t_s}\frac{dx_s}{dt}dt=\frac{d}{2}.$$

If $k(t)$ and dx_s/dt decrease with time, as is the case for the loading conditions under consideration, the average of $k(t)$ must be smaller than one. If therefore in Eq. (11) a constant value is to be taken for $k(t)$, this value must be smaller than one. An estimation of this constant can be obtained by solving the linear problem $B=\mu H$ in the interval $x_s\leq x\leq d/2$ with the condition that $E(x_s,t)=E_s=$constant, and finding the time t necessary for H to be H_s at the point x_s, and the increment Δt necessary for the average H in the interval $(x_s,x_s+\Delta x_s)$ to become H_s. This approach matches $E=E_s$ (and hence $\partial H/\partial x$), and H to their proper values at the point x_s, but it does not give for H its proper value in every point of the interval $(x_s,d/2)$. However, it gives a satisfactory solution if $H(0,t)$ is large compared to H_s, because for this case the main flux variation occurs in a small interval just to the right of x_s. The solution of the linear problem stated above results in the relation between dx_s and dt that is given by Eq. (12)

$$\frac{dx_s}{dt}=\frac{4E_s(t)}{\mu\pi H_s}. \tag{12}$$

The derivation of (12) is omitted since it involves only straightforward mathematics.

The following analysis will be based on Eq. (13)

$$dx_s/dt=E_s(t)/k\mu H_s \tag{13}$$

where k will be assumed to be a constant whose value as given by Eq. (12) is

$$k=\pi/4.$$

The form of $e(t)$ and the time necessary for the core to be saturated will be considered next.

4. Constant Driving Current

Let the source be a constant current driver so that $i(t)=I_0$; then $H(0,t)=H_0=N\cdot I_0/s$.

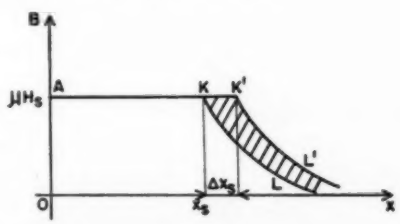


FIG. 5. Plot of B *versus* x; curve AKL for time t; curve $AK'L'$ for time $t+\Delta t$.

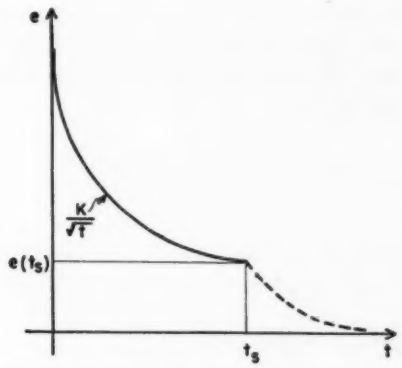


FIG. 6. Predicted voltage output during switching of a core that is driven by a constant current generator.

From Eqs. (9) and (13) we obtain the relation

$$\frac{dx_s}{dt}=\frac{H_0-H_s}{k\sigma\mu H_s x_s},$$

whose integral is given by Eq. (14) since $x_s(o)=0$

$$t=\frac{k\sigma\mu H_s}{2(H_0-H_s)}x_s^2. \tag{14}$$

In Eq. (14), t is the time necessary for the saturation point to be at x_s; the core is completely saturated when $x_s=d/2$; the corresponding time t_s is given by

$$t_s=\frac{k\sigma\mu H_s d^2}{8(H_0-H_s)}. \tag{15}$$

The time t_s will be called the saturation time; it is proportional to μH_s. The quantity μH_s equals the increase in flux density ΔB from the initial to the saturated state. Thus,

$$t_s=\frac{k\sigma\Delta B d^2}{8(H_0-H_s)}. \tag{15a}$$

To find the form of the induced voltage $e(t)$, we solve Eq. (14) for x_s and insert the result in (9). Using Eq. (6), we obtain Eq. (16).

$$e(t)=\frac{Nh}{t^{\frac{1}{2}}}\left(\frac{2k\mu H_s(H_0-H_s)}{\sigma}\right)^{\frac{1}{2}}. \tag{16}$$

The voltage $e(t)$ as given by Eq. (16) is plotted in Fig. 6. For $t=t_s$, $x=d/2$; hence (see Eq. (9))

$$e(t_s)=\frac{4Nh(H_0-H_s)}{\sigma d}.$$

The effect of μ_1 on the voltage wave form during the interval of time $0\leq t\leq t_s$ is negligible, it can be shown, if

$$\frac{\mu_1}{\mu}\ll\frac{H_s}{H_0-H_s}.$$

To find $e(t)$ for $t>t_s$, we must solve the linear problem $B=\mu_1 H$ with the boundary conditions set forth in Eqs. (4) and (5) and the initial condition at $t=t_s$ as given by Eq. (17).

$$H(x,t_s)=H_0-(H_0-H_s)\frac{2x}{d}. \tag{17}$$

Equation (17) is true because at $t=t_s$, H varies linearly from the value H_0 at $x=0$ to the value H_s at $x=d/2$.

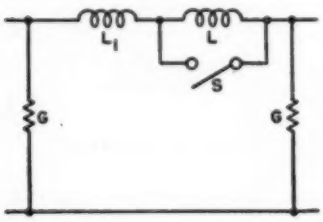


FIG. 7. One section of the filter of n sections that is approximately equivalent to the magnetic core. ($L+L_1)n=\mu d/2$; $nL_1=\mu_1 d/2$, $nG=\sigma d/4$.

An exact solution of the above linear problem can easily be obtained with known methods. The result is in the form of an infinite series. In the following an approximate simple solution will be given: Eqs. (1) and (2) have the same form as the voltage and current equation of a transmission line with $R=C=0$. The condition corresponding to Eq. (4) is obtained when a line of length $d/2$ is terminated with a short circuit. If the instantaneous current at any point in the line is i, the inductance per unit length is represented by a filter, a typical section will be as shown in Fig. 7, where the switch S closes when the current becomes i_s. It is of interest to note that if the line is represented by a filter with only one section, with $L+L_1=\mu d/2$, $L_1=\mu_1 d/2$, $G=\sigma d/2$, and the end shorted, the switching time t_s is obtained from the following expression for i_s:

$$i_s=I_0\left[1-\exp\left(-\frac{t_s}{LG}\right)\right]\cong\frac{I_0}{LG}t_s$$

(since $t_s/LG\ll 1$). Replacing the current in the expression for the equivalent filter by the magnetic field for

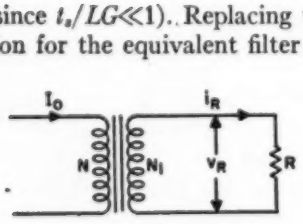


FIG. 8. Resistor loaded core, schematic.

the case of the magnetic core, we obtain for t_s the value $\mu\sigma d^2 H_s/8H_0$.

To find the voltage wave form for $t>t_s$ we use the equivalent circuit of Fig. 7 with the switch S closed. Letting $t'=t-t_s$, we have as initial current in the coil the average value $(H_0+H_s)/2$ of the field inside the core at time $t'=0$; thus it is easily shown that $E(0,t)$ (corresponding to the voltage in the input of the filter) is given by the relation

$$E(0,t)=\frac{2(H_0-H_s)}{\sigma d}\exp\left(-\frac{8t'}{\mu_1\sigma d}\right).$$

Hence,

$$e(t')=2hE(0,t)=\frac{4hN(H_0-H_s)}{\sigma d}\exp\left(-\frac{8t'}{\mu_1\sigma d}\right)$$

$$=e(t_s)\exp\left(-\frac{8t'}{\mu_1\sigma d}\right). \tag{18}$$

The voltage form $e(t')$, for $t>t_s$ is thus an exponential with a time-constant T given by

$$T=\mu_1\sigma d/8.$$

The dotted curve in Fig. 6 indicates the voltage wave form for $t>t_s$.

III. CORE UNDER LOAD

1. Resistive Load

Again assume that the driver is a constant current generator. If a resistive load, R, is connected to the core, we have (see Fig. 8)

$$H(0,t)=\frac{NI_0}{s}-\frac{N_1 i_R}{s}$$

and

$$v_R=Ri_R=2hN_1E(0,t)$$

[see Eq. (9)]. Hence,

$$H(0,t)=H_0-GE(0,t), \tag{19}$$

where

$$H_0=NI_0/s$$

and

$$G=\frac{2h}{sR}N_1^2. \tag{20}$$

From Eqs. (9) and (19) we obtain Eq. (21):

$$E_s(t)=\frac{H_0-H_s}{\sigma x_s+G}. \tag{21}$$

Hence, with Eq. (13),

$$\frac{dx_s}{dt}=\frac{H_0-H_s}{k\mu(\sigma x_s+G)}. \tag{22}$$

Integrating Eq. (22), we get

$$t=\frac{k\sigma\mu H_s}{H_0-H_s}\left(\frac{x_s^2}{2}+\frac{G}{\sigma}x_s\right). \tag{23}$$

The saturation time t_s is given by Eq. (23a) which

follows from (23) when $x_s=d/2$.

$$t_s=\frac{k\mu\sigma H_s d^2}{8(H_0-H_s)}\left(1+\frac{4G}{\sigma d}\right)=t_{s0}\left(1+\frac{4G}{\sigma d}\right). \quad (23a)$$

In Eq. (23a), t_{s0} is the saturation time without load. To find the form of $e(t)$ we solve Eq. (23) for x_s and insert the result in Eq. (21). Knowing $E(0,t)$, we find $e(t)$ from Eq. (6). The result is plotted in Fig. 9 for three cases: $G=0$, $d/4$, $d/2$.

For $t=t_s$,

$$e(t_s)=\frac{2Nh(H_0-H_s)}{G+(\sigma d/2)}. \quad (24)$$

For G large compared to $\sigma d/2$,

$$E_s(t)\simeq\frac{H_0-H_s}{G}=\text{constant};$$

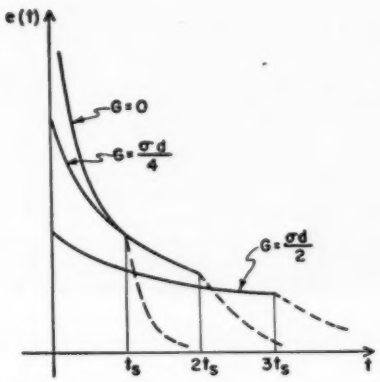


FIG. 9. Predicted voltage output during switching of a resistance loaded core for three different values of the load resistance.

hence in this case the voltage wave form during switching remains approximately constant.

For $t>t_s$, an approximate expression for $e(t)$ can be obtained, as in the case of the core without load, by using the one-section equivalent filter. The equivalent load, G, will appear as an input load on the filter. The resulting voltage wave form is an exponential curve with a time constant.

$$T=\frac{\mu_1\left[G+\frac{\sigma d}{4}\right]}{2}$$

and is shown by the dotted lines in Fig. 9.

2. Capacitive Load

With a constant driving current I_0 and a capacitive load C, we have (see Fig. 10)

$$H(0,t)=\frac{NI_0}{s}-\frac{N_1 i_c}{s}$$

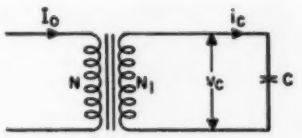


FIG. 10. Capacitor loaded core, schematic.

and

$$\frac{dv_c}{dt}=\frac{i_c(t)}{C}=2hN_1\frac{dE(0,t)}{dt}.$$

Hence,

$$H(0,t)=H_0-C'\frac{dE(0,t)}{dt}, \quad (25)$$

where

$$C'=\frac{2h}{s}N_1^2C.$$

From Eqs. (9) and (24), we obtain Eq. (26).

$$H_0-C'\frac{dE_s}{dt}-H_s=\sigma x_s E_s(t). \quad (26)$$

With Eq. (13), Eq. (26) leads to

$$\rho\mu kC'\frac{d^2x_s}{dt^2}+\rho\mu kx_s\frac{dx_s}{dt}=1, \quad (27)$$

where

$$\rho=H_s/(H_0-H_s).$$

The first integration of Eq. (27) gives Eq. (28),

$$\rho\mu kC'\frac{dx_s}{dt}+\frac{\rho\mu k}{2}x_s^2=t, \quad (28)$$

which is a special form of Riccati's equation. Figures 11 and 12, respectively, give plots of $x_s(t)$ and $dx_s(t)/dt$ for different values of C. The induces voltage $e(t)$ has the form of Fig. 12 since it is proportional to dx_s/dt (see Eqs. (6) and (13)). For large values of C,

$$dx_s/dt\simeq t/k\mu C';$$

hence, in this case, the voltage wave form during switching varies linearly with time.

For $t>t_s$, the voltage form is the same as for the equivalent transmission line with a capacitor across its input, in parallel with a constant current generator. For an approximate expression we use the one-section equivalent filter with a capacity C' in the input. Since

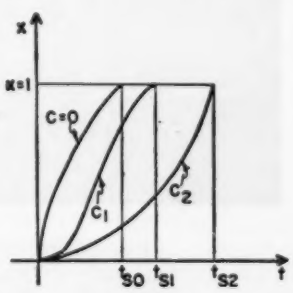


FIG. 11. The saturation point plotted *versus* t for a capacitor loaded core, for three different values of the load capacity.

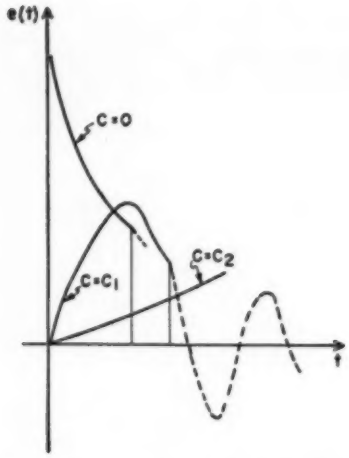


FIG. 12. Predicted voltage output during switching of a capacitor loaded core for three different values of the load capacity.

oscillations are possible, we must consider the return path of the B-H curve; assume that it remains a straight line with slope μ_1. The resulting wave form is given by the dotted line in Fig. 12 for a value of C for which oscillations occur.

TABLE I.

Core number	H_1 oersted	B gauss	h inches	r inches	d mils	mh o/m	w Number of wraps
1	0.13	12 000	$\frac{1}{4}$	$\frac{3}{16}$	1	$10^{-6}/40$	8
2	0.11	10 500	$\frac{1}{4}$	$\frac{3}{16}$	2	$10^{-6}/40$	4
3	0.10	10 000	$\frac{1}{4}$	$\frac{3}{16}$	4	$10^{-6}/40$	2

3. The Switching Time for an Arbitrary Magnetizing Force

It will be shown that when the magnetizing force is an arbitrary function of time $H_0(t)$, the saturation time is given by the same expression as in Eq. (15a) where

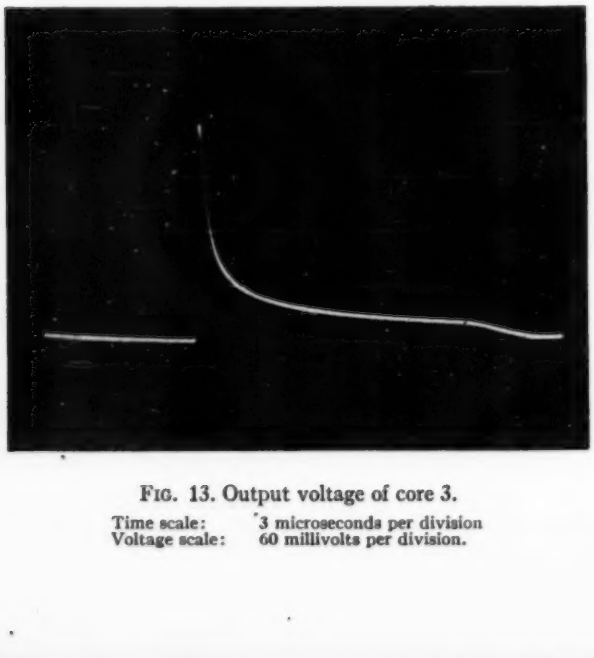


FIG. 13. Output voltage of core 3.
Time scale: 3 microseconds per division
Voltage scale: 60 millivolts per division.

H_0 must be replaced by the average value $\bar{H}_0$ of $H(t)$. The average value is defined in Eq. (29).

$$\bar{H}_0=\frac{1}{t_s}\int_0^{t_s} H_0(t)dt. \tag{29}$$

From Eqs. (11a) and (13), we have

$$\frac{dx_s}{dt}=\frac{H_0(t)-H_s}{k\sigma\mu H_s x_s},$$

which, when integrated, gives

$$\frac{k\sigma\mu H_s}{2}x_s^2=\int_0^t [H_0(t)-H_s]dt.$$

For $x_s=d/2$, $t=t_s$. Thus, replacing in the foregoing,

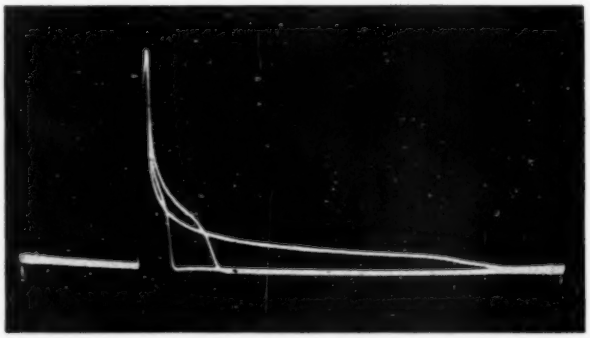


FIG. 14. Output voltage of cores 1, 2, and 3.
Time scale: 3 microseconds per division
Voltage scale: for core 1 230 millivolts per division
for core 2 130 millivolts per division
for core 3 60 millivolts per division.

we obtain

$$\frac{k\sigma\mu H_s d^2}{8}=\int_0^{t_s} H_0(t)dt-H_s t_s=(\bar{H}_0-H_s)t_s.$$

Therefore,

$$t_s=\frac{k\sigma\mu H_s d^2}{8(\bar{H}_0-H_s)}=\frac{k\sigma\Delta B d^2}{8(\bar{H}_0-H_s)},$$

which has the same form as Eq. (15a). The foregoing is true if $H_0(t)$ remains large compared to H_s.

IV. EXPERIMENTAL RESULTS

Measurements were performed with three magnetic cores. The core material was 4–79 Permalloy; the physical properties of the cores are listed in Table I. The driver was a constant current generator with rise time of about 0.5 microsecond. A single turn served as the input winding.

Figure 13 shows the voltage output of core 3 (Table I) without load; driving current I_0, is 6 amperes, corresponding to $H_0=202$ ampere-turns per meter; the

read-out coil consisted of one turn. Figure 13 has the same form as Fig. 6. The theoretical value of $e(t)$ is obtained from Eq. (16) by multiplying the number of wraps w. Thus, from theory

$$e(t)=69\times10^{-6}\times\frac{1}{t^{\frac{1}{2}}}\text{ volts},$$

which is in close agreement with Fig. 13. The saturation time computed from Eq. (15), is $t_s=13$ microseconds.

Figure 14 gives the voltage outputs of cores 1, 2, and 3 (Table I) under the same conditions as already given; for cores land 2 the theoretical expressions for $e(t)$ are

$$\text{core 1:}\quad e(t)=276\times10^{-6}\times\frac{1}{t^{\frac{1}{2}}}\text{ volts},$$

$$\text{core 2:}\quad e(t)=138\times10^{-6}\times\frac{1}{t^{\frac{1}{2}}}\text{ volts},$$

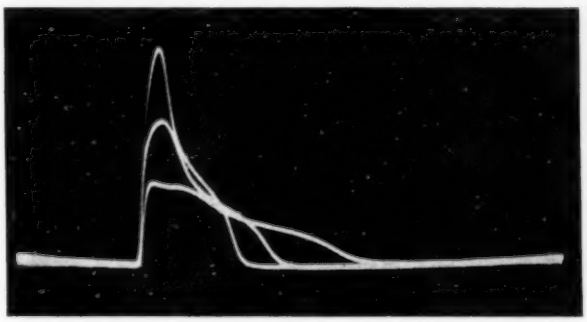


FIG. 15. Output voltage of core 1 for three values of applied magnetizing field.

Driving current: I_0=6A, 4.5A and 3A
Time scale: 1 microsecond per division
Voltage scale: 220 millivolts per division.

which is in agreement with the voltage wave forms in Fig. 14.

The corresponding saturation times are

$$\text{core 1:}\quad t_s=0.8\text{ microseconds},$$

$$\text{core 2:}\quad t_s=3.25\text{ microseconds}.$$

Figure 15 shows the voltage output of core 1 without load for three different values of applied magnetizing field. The three current values $I_0=6$ amperes, 4.5 amperes, 3 amperes, correspond to $H_0=202\ A-t/M$, 151 $A-t/M$, 101 $A-t/M$.

Figure 16 shows the voltage output of core 1 for three different conditions: without load, with a load resistance R, of 94 ohms, and with a load resistance R, of 42 ohms. The output winding N_1 contains 30 turns, and the driving current I_0 is 6 amperes. The voltage was observed across the load resistor. From Eq. (23a), we have

$$t_s=t_{s0}[1+(4G/\sigma d)],$$

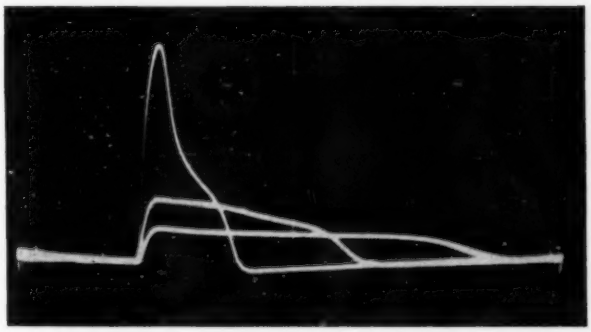


FIG. 16. Output voltage of core 1 for three values of load resistor.

Time scale: 1 microsecond per division
Voltage scale: 6.9 volts per division.

where

$$G=(2h/sR)N_1^2w$$

an obvious modification of Eq. (23) for the case of more than one wrap of material in the core.

$$\text{For } R=94:\quad 1+(4G/\sigma d)=2;\ t_s=2t_{s0}.$$

$$\text{For } R=42:\quad 1+(4G/\sigma d)=3.3;\ t_s=3.3t_{s0},$$

where t_{s0} is the saturation time without load. At $t=0$,

$$e(0)=\frac{2N_1h(H_0-H_s)}{G}\,w\text{ volts}$$

at $t=t_s$

$$e(t_s)=\frac{2N_1h(H_0-H_s)}{G+\sigma d/2}\,w\text{ volts}$$

$$\text{For } R=94:\quad e(0)=9.5\text{ volts};\ e(t_s)=3.2\text{ volts}$$

$$\text{For } R=42:\quad e(0)=4.2\text{ volts};\ e(t_s)=2.3\text{ volts}.$$

The theoretical values above agree closely with the values measured in Fig. 16; the form of the curves in Fig. 16 is the same as in Fig. 9.

Figure 17 gives the output of core 1 for three different

FIG. 17. Output voltage of core 1 for three values of loading capacitance.

Time scale: 1 second per division
Voltage scale: 6.9 volts per division.

conditions: without load; with capacitive loads, $C=0.015$ microfarad; and $C=0.085$ microfarad. The wave forms in Fig. 17 have the same form as the curves in Fig. 12.

V. CONCLUSIONS

The results that are developed in Section II, although approximate, describe satisfactorily the performance of the core for the case of high magnetizing force, as can be seen from the close agreement between the theoretical and the experimental wave forms. Secondary factors like variation of μ inside the material, and the rise time of the constant current source have been omitted. The secondary factors, however, affect the voltage wave form even for high H when the thickness of the core material is less than $\frac{1}{2}$ mil.

For H of the order of H_1 (see Fig. 1), the foregoing treatment does not hold.[2]

ACKNOWLEDGMENTS

This work was done in the Magnetics Section of the Burroughs Research Center in Philadelphia.

The paper was developed after discussions with Mr. John O. Paivinen and Mr. Joseph Wylen. Their many suggestions contributed greatly to its preparation. Mr. Wylen furnished the experimental results and assisted in the preparation of the paper. The Magnetic Metals Company, Camden, New Jersey, supplied the experimental magnetic cores.

[2] The following general statement for an arbitrary concave B-H curve has been proved: if the magnetizing force is constant the output voltage waveform is a decreasing function of time.

JOURNAL OF APPLIED PHYSICS VOLUME 25, NUMBER 2 FEBRUARY, 1954

Ferromagnetic Spinels with Rectangular Hysteresis Loops

I. J. Hegyi
Radio Corporation of America, RCA Laboratories Division, Princeton, New Jersey
(Received May 25, 1953)

The ferromagnetic spinel $MgO \cdot 3MnFe_2O_4$, "magnesium manganese ferrite" is satisfactory for magnetic memory devices. Small rings of about 55-mil diameter which have proved practical for memory matrix use have a coercive force of about 3 oersteds and switch-over time of 0.3 microsecond, and a nearly rectangular hysteresis loop. Rings of about 0.5-cm diameter which were used to study composition variables showed a lower coercive force of 1.6 oersteds for the same composition and processing. Resistivities are in the order of 45×10^3 ohm cm.

1. INTRODUCTION

In magnetic memory devices using arrays of magnetic cores, such as worked on at RCA Laboratories by Rajchman's group,[1] and others,[2–4] it is advantageous to have a magnetic material which has a nearly rectangular hysteresis loop, a low coercive force and a high resistance. The rectangular hysteresis loop is necessary, because such materials have two stable states of positive and negative residual flux density to store binary information. The low coercive force is advantageous to minimize the current required to store information.

Metallic magnetic materials have been available which meet the rectangularity and low coercive force requirements. These materials are cold-rolled nickel-iron alloys and oriented silicon steels, although such materials have a long switch-over time of about 5 to 10 000 microseconds, and they are rather expensive.

Ferromagnetic spinel materials with their inherent high resistivity and consequent low eddy currents have a switch-over time of a few tenths of a microsecond and are much less costly.

Preliminary tests showed that the hysteresis loop of the magnetic spinel $MnFe_2O_4$ approaches rectangularity. Also, it appeared that if the loop rectangularity of this material could be increased without greatly increasing the coercive force, a useful material could be obtained. Such was found to be the case, and this paper describes the preparation and properties of an altered manganese ferrospinel which has a rectangular hysteresis loop, a relatively low coercive force, and a high resistivity.

2. EXPERIMENTAL

(a) Test Procedure

In order to evaluate the materials as to their degree of rectangularity of hysteresis loop, it has been found desirable to make tests under conditions approximating those to be found in practical application as an element of the memory array. The test procedure used was a current-pulse method which was developed and constructed by Rajchman and Arbaeus, and all tests by this method were made by Arbaeus. The test sample was prepared as a ring having an outside diameter of about 0.5 cm, an inside diameter of about 0.3 cm, and a height of about 0.15 cm. The ring was wound as a transformer through the primary of which flowed rec-

[1] J. A. Rajchman, RCA Rev. **13**, 183–201 (1952).
[2] A. Wang, J. Appl. Phys. **21**, 49–54 (1950).
[3] J. W. Forrester, J. Appl. Phys. **22**, 44–48 (1951).
[4] W. N. Papian, Proc. Inst. Radio Engrs. **40**, 475–478 (1952).

tangular current pulses. These current pulses as a series of five in number had the following polarities and amplitudes: pulse 1, positive and unit amplitude; pulses 2 and 3 negative, one-half amplitude; pulse 4 positive and unity; pulse 5 negative and unity. By comparing the voltages induced by these pulses in the secondary winding, certain points on the hysteresis loop, as illustrated in Fig. 1 may be determined. In Fig. 1, point 6 represents the magnetic induction remaining in a test ring immediately after completion of pulses 1 and 4, point 7 represents the magnetic induction remaining immediately after completion of pulse 2, point 8 represents the magnetic induction remaining immediately after completion of pulse 3, and point 9 represents the magnetic induction remaining immediately after completion of pulse 5. The parameter of the rectangularity is defined as

$$D=B_R/(B_R-B_D),$$

where D is called the discrimination, B_R is the magnetic

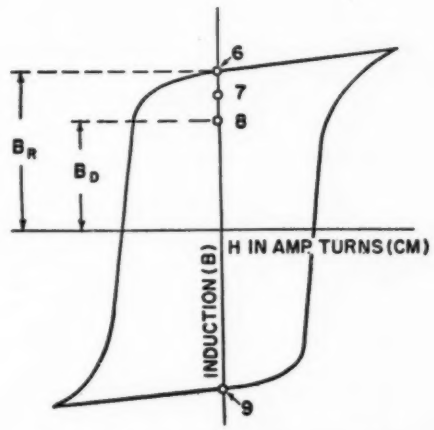


FIG. 1. Points on hysteresis loop by current pulse method. Point 6, induction after pulses 1 and 4. Point 7, induction after pulse 2. Point 8, induction after pulse 3. Point 9, induction after pulse 5.

induction at point 6 of Fig. 1, and B_D is the magnetic induction at point 8 of the figure. It may be seen that the value of D varies directly with the degree of rectangularity. The curve on the trailing edge of the hysteresis loop becomes sharper as B_R increases and as (B_R-B_D) decreases. For magnetic memory devices it is also desirable that the material have a low coercive force. In order to evaluate the material, including both its coercive force and the rectangularity of its hysteresis characteristic, a quality factor chosen as D/H was used, where H is the field at maximum discrimination. The value of H in ampere turns/cm is obtained by the relation

$$H=IN/1,$$

where I is the current at maximum discrimination, N is the number of turns on the primary winding of the test ring, and 1 is the average magnetic path in cm of the test ring. The value H is not the coercive force of the material but is a satisfactory measurement for obtaining the quality factor D/H which gives a relative evaluation of the discrimination and coercive force of a material.

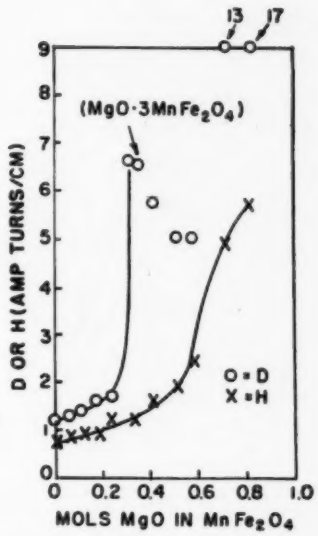


FIG. 2. D and H of MgO plus $MnFe_2O_4$ in solid solution.

A few of the compositions were tested on a calibrated 60-cycle/second dynamic hysteresis-loop tracer constructed by H. L. Donley. With this apparatus, the saturation value B_S, the remanence B_R, and the coercive force H_C were obtained.

The initial permeability μ_0 and Q were measured at 50 kc on a rf bridge.

(b) Material Preparation

The materials were synthesized from their constituent oxides MgO, MnO_2, and Fe_2O_3 by ceramic processing methods already described.[5,6] Manganese carbonate

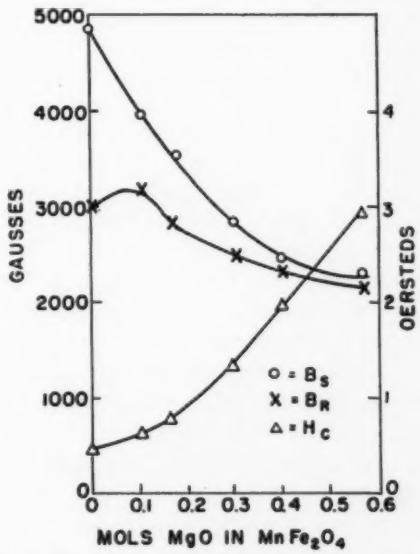


FIG. 3. B_S, B_R, and H_C of MgO plus $MnFe_2O_4$ in solid solution.

[5] J. L. Snoek, *New Developments in Ferromagnetic Materials* (Elsevier Publishing Company, Inc., New York, 1947).
[6] Harvey, Hegyi, and Leverenz, RCA Rev. 11, 321–363 (1950).

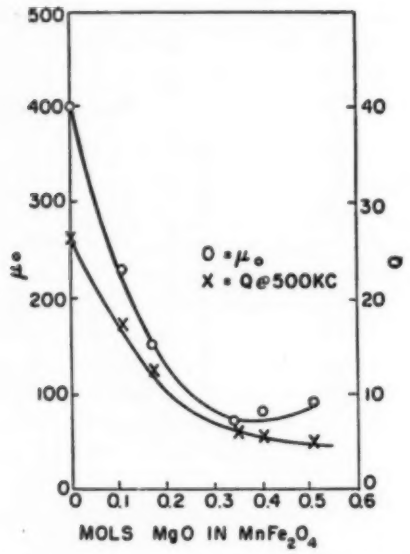


FIG. 4. μ_0 and Q at 500 kc of MgO plus $MnFe_2O_4$ in solid solution.

$MnCO_3$ may also be used with equivalent results, and in some cases may be preferable, as not all grades of MnO_2 give satisfactory results.

The materials whose characteristics are plotted in Fig. 2 were ring samples of 0.5-cm diameter. The plot shows that the loop-rectangularity of $MnFe_2O_4$ is increased with addition of MgO incorporated in solid solution, with a corresponding increase in H. At compositions having MgO greater than about 0.33 MgO, which is equivalent to $MgO \cdot 3MnFe_2O_4$, the values of D become sensitive to variations in heating procedure, and their reproducibility becomes more difficult with further addition of MgO. As H values above about 2.0 ampere turns/cm are not desirable, the best material for memory-device use is $MgO \cdot 3MnFe_2O_4$, using D/H as the criterion of quality. The values of D and H vary with the temperature and the length of time of crystallization.

Data were obtained, also, on remanence, saturation, and coercive-force values by using a calibrated 60-cycle/second dynamic loop tracer; and initial permeability, and Q at 500 kc were obtained on a rf bridge. For these measurements, ring samples of 2.54-cm outside diameter were used. Figure 3 shows data on these materials as obtained from the hysteresis-loop traces. The coercive force H_C increases as B_S and B_R approach each other with increasing amounts of MgO. Figure 4 shows the decrease of initial permeability, μ_0 and Q with increasing amounts of MgO.

The Curie temperatures of the materials were found to decrease linearly from 320°C, for $MnFe_2O_4$, to 300°C, for $MgO \cdot 3MnFe_2O_4$. The lattice parameter, or cube edge, also decreases linearly from 8.51A for $MnFe_2O_4$ to 8.46A for $MgO \cdot 3MnFe_2O_4$. X-ray powder-diffraction patterns showed a solid solution of MgO in $MnFe_2O_4$ for all samples. The resistivity of the $MgO \cdot 3MnFe_2O_4$ was 45×10^3 ohm cm.

Solid solutions of $MgFe_2O_4$ in $MnFe_2O_4$ were prepared also by heating mixed oxides in proper proportion. These materials were tested by the current-pulse method and are listed in Table I.

3. DISCUSSION AND CONCLUSION

The experimental data presented show that the rectangularity of the hysteresis loop of $MnFe_2O_4$ is increased if MgO is added in solid solution. An increase in coercive force is also obtained. This, however, is not appreciable for the above combinations, and so the quality factor D/H is appreciably increased. In the case of solid solution of $MgFe_2O_4$ in $MnFe_2O_4$ as shown in Table I, there is a large increase in H, as D increases, and so no improvement in D/H results.

The material $MgO \cdot 3MnFe_2O_4$ showed the most promise for magnetic-memory use, because this material is reproducible, has a relatively low coercive force, and large discrimination. If the MgO content is further increased, the reproducibility is reduced. This reduction is caused, most likely, by the known characteristic of magnesium ferrospinels where the magnesiums take positions in either four or six coordinated positions in the close-packed oxygen lattice.[7]

TABLE I.

Molar proportions $MgFe_2O_4$	Molar proportions $MnFe_2O_4$	D	H	D/H
0.25	0.75	1.9	1.4	1.35
0.50	0.50	3.4	4.1	0.83
0.60	0.40	5.7	5.3	1.07
0.66	0.34	6.0	6.1	0.98
0.90	0.10	6.6	5.4	1.22

For use in a memory magnetic matrix, rings having an outside diameter of about 55 mils, inside diameter of 35 mils, and height of 15 mils were made from the $MgO \cdot 3MnFe_2O_4$ material. At optimum discrimination these rings had a switch-over time of 0.3 microsecond when driven with current of 1.1 ampere turns. This maximum drive corresponds to 3.9 oersteds which is equivalent to a coercive force of about 3 oersteds. These smaller rings at present have proved more practical for memory matrix use than the 0.5-cm outside diameter rings for which data is given in this paper. The experiments have indicated that using the $MgO \cdot 3MnFe_2O_4$ composition and same processing larger diameter rings (0.5 cm) have a smaller coercive force (1.6 oersteds).

[7] L. Néel, Ann. phys. 3, 137 (1948).

JOURNAL OF APPLIED PHYSICS VOLUME 25, NUMBER 2 FEBRUARY, 1954

Coupling of Modes of Propagation

J. R. PIERCE
Bell Telephone Laboratories, Murray Hill, New Jersey
(Received June 8, 1953)

When two lossless modes of propagation are coupled, waves which increase or decrease with distance may arise when the power flow of the two modes is in opposite directions or when power is generated in the coupling device. This behavior is characteristic of wave filters, traveling-wave tubes, double-stream amplifiers, and space-charge-wave amplifiers. Such behavior is analyzed assuming linearity and conservation of energy only.

INTRODUCTION

MANY important devices and effects involve coupling between modes of propagation. For instance, a wave filter may be made by placing a series of obstacles in a waveguide. The obstacles couple the forward and backward modes of propagation and give rise to pass bands and stop bands. A series of loading coils in a transmission line and a series of weights attached to a stretched string have similar effects. Directional couplers are another example of devices utilizing the coupling of modes of propagation.

In electron devices, we can regard the gaining wave in a traveling-wave tube as arising through coupling between a forward-moving circuit wave and the slower space-charge wave, which moves backward relative to the electrons of the beam. Double-stream amplification can be regarded as arising through coupling between two space-charge waves. Space-charge-wave amplification can be considered as operation in the stop band of a wave filter in which the space-charge waves are reflected at periodic discontinuities.

Although it is easy to make broad general analogies and to draw broad conclusions, it is also important to be able to check these, to prove theorems and to make calculations. One method of achieving rigorous and quantitative results is to formulate the equations governing a particular case of mechanical waves, or electromagnetic waves, or space-charge waves, and to work with these equations, as W. E. Mathews has done in the case of transmission lines and electron streams in relative motion.[1,2] However, it is not always easy to formulate equations for a particular system with a satisfying degree of generality. Further, in dealing with particular instances one always wonders just how broad analogies are.

Another approach is to abstract from the various sorts of systems which are to be considered a few very elementary properties which they have or which they may have in common, and to try to deduce various forms of common behavior from these elementary properties.

This paper attempts to do this for the very simple case of coupling between two lossless, unattenuated modes of propagation of a linear system. The treatment is based chiefly on linearity and conservation of energy, although reciprocity is considered briefly.

It is quite possible that the approach used here could be used to investigate cases of coupling between more than two modes of propagation. Counting forward and backward propagation in each guide as two modes, a directional coupler has four modes. Thus, the work given here does not apply strictly to directional couplers. However, the writer felt it best to tackle a simple case first. This is perhaps justified by the fact that this simple case, treated by familiar methods, can serve to relate and explain the operation of a variety of devices.

1. LINEAR TRANSDUCER

Consider a linear system with two unattenuated modes of propagation. Let the two modes be uncoupled except in a region of limited extent, and there let them be coupled in a linear manner.

Let us measure the complex amplitudes of the two modes by P and Q. P and Q are proportional to some quantity, such as voltage or current or displacement or velocity in terms of which power flow can be expressed, and P and Q are chosen so that the power flow in the first mode is PP^* and the power flow in the second mode is $\pm QQ^*$. Choice of the plus sign means that power flow is in the same direction for both modes; choice of the minus sign means that the two modes carry power in opposite directions.

The situation to be considered is represented in Fig. 1. Here P_a is the amplitude of one mode at one end of the region of coupling, P_d that at the other end of the region of coupling, and similarly for Q_a and Q_d. $r_1 \cdots r_4$ are reference planes at which the complex amplitudes P_a, P_d, Q_a, Q_d are measured. We will call the region of coupling a transducer.

The system considered has the same properties as the usual "four terminal network," and if we regard such a network as being connected to transmission lines at each

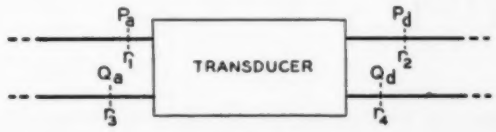


FIG. 1.

[1] W. E. Mathews, J. Appl. Phys. **22**, 310–316 (1951).
[2] W. E. Mathews, Proc. Inst. Radio Engrs. **39**, 1044–1051 (1951).

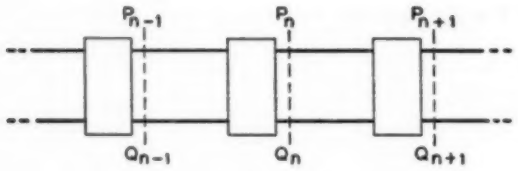


FIG. 2.

end, P_a and Q_a can be regarded as measuring the amplitudes of two waves at one end of the network and P_d and Q_d as measuring the amplitudes of two waves at the other end of the network.

Because the modes and the transducer are assumed to be linear, we can write

$$P_d = AP_a + BQ_a, \tag{1.1}$$

$$Q_d = CQ_a + DP_a. \tag{1.2}$$

Ordinarily, modes are considered in pairs, as the forward and the backward waves on a transmission line. This pair of modes has been treated, for instance, by Fano and Lawson.[3] Here we will proceed with no assumption as to the nature of the modes or their velocities.

2. A CHAIN OF LINEAR TRANSDUCERS

Suppose that the two modes are coupled periodically, as illustrated in Fig. 2. We see that P_n and P_{n+1} correspond respectively to P_a and P_d of (1.1), and also Q_n and Q_{n+1} correspond to Q_a and Q_d of (1.2).

Let us assume that mode P suffers a phase lag θ_p between transducers and mode Q suffers a phase lag θ_q between transducers. Using (1.1) and (1.2) we write

$$P_{n+1} = Ae^{-j\theta_p}P_n + Be^{-j\theta_q}Q_n, \tag{2.1}$$

$$Q_{n+1} = Ce^{-j\theta_q}Q_n + De^{-j\theta_p}P_n. \tag{2.2}$$

We can solve (2.1) for Q and substitute Q_n and Q_{n+1} in (2.2). This gives

$$P_{n+2} - (Ae^{-j\theta_p} + Ce^{-j\theta_q})P_{n+1} + (AC - BD)e^{-j(\theta_p+\theta_q)}P_n = 0. \tag{2.3}$$

Let us look for solutions in the form

$$P_{n+1} = P_n e^{j\beta} = m \exp\left[-j\frac{(\theta_p+\theta_q)}{2}\right]P_n. \tag{2.4}$$

We obtain an equation in m

$$m^2 - \left\{A \exp\left[j\frac{(\theta_q-\theta_p)}{2}\right] + C \exp\left[-j\frac{(\theta_q-\theta_p)}{2}\right]\right\}m + (AC - BD) = 0. \tag{2.5}$$

[3] *Microwave Transmission Circuits*, edited by George L. Ragan, (McGraw-Hill Book Company, Inc., New York, 1948), Chap. 9, Sec. 9.3.

3. CONSERVATION OF ENERGY

Let us now assume that the two modes in question travel in a fixed medium. There is no source of energy, and the transducer is assumed to be lossless. Thus, the power flowing into the transducer must be equal to the power flowing out, or

$$P_aP_a^* \pm Q_aQ_a^* = P_dP_d^* \pm Q_dQ_d^*. \tag{3.1}$$

If we use (1.1) and (1.2) in connection with (3.1), we obtain

$$P_aP_a^*(AA^* \pm DD^* - 1) + Q_aQ_a^*(BB^* \pm CC^* \mp 1) + [P_aQ_a^*(AB^* \pm DC^*)] + [P_aQ_a^*(AB^* \pm DC^*)]^* = 0. \tag{3.2}$$

Now (3.2) must hold for $P_a = 0$ and for $Q_a = 0$. Hence, we must have

$$AA^* \pm DD^* - 1 = 0, \tag{3.3}$$

$$BB^* \pm CC^* \mp 1 = 0. \tag{3.4}$$

If this is so, the last line says that

$$R_e(P_aQ_a^*)(AB^* \pm DC^*) = 0. \tag{3.5}$$

The phase of $P_aQ_a^*$ is arbitrary, so that (3.5) means that

$$AB^* \pm DC^* = 0. \tag{3.6}$$

So far, we have allowed an arbitrary location of the reference planes $r_1 - r_4$ at which the P's and Q's are measured. If $Q_a = 0$, for any location of r_1 we can locate r_2 and r_4 so that P_d and Q_d are in phase with P_a. Thus, for any location of r_1 we can shift r_2 and r_4 so that A and D are real. Suppose we do this. Then if $P_a = 0$, we can locate r_3 so that C is real. But we see through (3.6) that if A, D and C are real, B must be real. Thus, for any location of r_1, A, B, C, D can be made real by locating r_2, r_3 and r_4 properly. Suppose this has been done.

We find that (3.3)–(3.5) are satisfied when A, B, C and D have real values such that (1.1) and (1.2) become

$$P_d = (1 \mp k^2)^{\frac{1}{2}}P_a + kQ_a, \tag{3.7}$$

$$Q_d = (1 \mp k^2)^{\frac{1}{2}}Q_a \mp kP_a. \tag{3.8}$$

Here the upper signs apply when the powers flow in the same direction and the lower signs when the powers flow in opposite directions. The degree of coupling is measured by the magnitude of k. Equations (3.7) and (3.8) apply to a general lossless linear transducer with specially chosen reference planes.

Suppose we wish to shift the reference planes to some other locations, so as to obtain the most general relation from (3.7) and (3.8). If we let P_a, P_d, Q_a, Q_d be shifted in phase by angles θ_I, θ_{II}, θ_{III}, and θ_{IV}, respectively, we obtain

$$P_d = (1 \mp k^2)^{\frac{1}{2}}e^{j(\theta_I - \theta_{II})}P_a + ke^{j(\theta_{III} - \theta_{II})}Q_a, \tag{3.9}$$

$$Q_d = (1 \mp k^2)^{\frac{1}{2}}e^{j(\theta_{III} - \theta_{IV})}Q_a \mp ke^{j(\theta_I - \theta_{IV})}P_a. \tag{3.10}$$

If we let

$$\theta_I-\theta_{II}=\theta_1, \tag{3.11}$$

$$\theta_{III}-\theta_{II}=\theta_2, \tag{3.12}$$

$$\theta_{III}-\theta_{IV}=\theta_3, \tag{3.13}$$

then

$$\theta_I-\theta_{IV}=\theta_1-\theta_2+\theta_3 \tag{3.14}$$

and

$$P_d=(1\mp k^2)^{\frac{1}{2}}e^{j\theta_1}P_a+ke^{j\theta_2}Q_a, \tag{3.15}$$

$$Q_d=(1\mp k^2)^{\frac{1}{2}}e^{j\theta_3}Q_a\mp ke^{j(\theta_1-\theta_2+\theta_3)}P_a. \tag{3.16}$$

We have in (3.15) and (3.16) the most general possible relation between P_a, P_d, Q_a, Q_d for lossless modes consistent with linearity and conservation of energy.

Let us use the values of A, B, C, D given in (3.15) and (3.16) in connection with the relation (2.5) for propagation in a chain of transducers. We obtain

$$m^2+(1\mp k^2)^{\frac{1}{2}}\left\{\exp\left[j\left(\frac{\theta_q-\theta_p}{2}+\theta_1\right)\right]+\exp\left[-j\left(\frac{\theta_q-\theta_p}{2}-\theta_3\right)\right]\right\}m+e^{j(\theta_3+\theta_1)}=0. \tag{3.17}$$

Let

$$M=m\exp\left[-j\left(\frac{\theta_3+\theta_1}{2}\right)\right]. \tag{3.18}$$

From (3.21) and (3.20) we obtain

$$M^2+2M(1\mp k^2)^{\frac{1}{2}}\cos\left(\frac{\theta_q-\theta_p+\theta_1-\theta_3}{2}\right)+1=0. \tag{3.19}$$

If the power flow is in the same direction for both modes, the minus sign applies and the radical is less than unity. If this is so, the equation always yields complex values of M of magnitude unity; that is, unattenuated waves. This is as it should be. If two lossless modes with power flow in the same direction are coupled together, power can flow back and forth between them but it must continue to travel along undiminished. In our picture, two new unattenuated modes are brought into being.

If, however, the power flow is in the opposite direction for the two modes, the plus sign applies and the radical is greater than unity. When this is so, (3.19) yields unattenuated waves only when

$$(1+k^2)^{\frac{1}{2}}\cos\left(\frac{\theta_q-\theta_p+\theta_1-\theta_3}{2}\right)<1. \tag{3.20}$$

Over the range of the argument of the cosine for which

$$(1+k^2)^{\frac{1}{2}}\cos\left(\frac{\theta_q-\theta_p+\theta_1-\theta_3}{2}\right)>1, \tag{3.21}$$

M is real, the waves described by M change in amplitude exponentially with distance, and the behavior is in all respects that characteristic of the stop band of a filter.

In the stop band, the net power flow is zero. Power flowing in by one mode is completely carried back by the other. Thus in terms of our variables P and Q, in a stop band $P_nP_n^*$ is equal to $Q_nQ_n^*$.

The parameter k measures the degree of coupling. For zero coupling, $k=0$, and the waves are always unattenuated. As the coupling is made larger, k becomes larger and the range of the argument of the cosine over which the wave decays exponentially becomes larger. As k is increased the stop band becomes broader with respect to the parameters governing the arguments of the cosine. In actual cases, θ_p, θ_q, θ_1, and θ_2 may vary with frequency, with electron velocity, or with other parameters.

Let us consider the case in which the transducer involves negligible phase shift, so that we can regard θ_1 and θ_3 as equal to zero. Then the stop bands are centered about

$$\theta_p-\theta_q=2n\pi. \tag{3.28}$$

An example of this is the stop band of a series of small regularly spaced obstacles in a transmission line when the obstacles are an integral number of half-wavelengths apart.

4. RECIPROCITY

So far we have invoked linearity and the conservation of energy, but we have said nothing about reciprocity. Let us consider it for the sake of completeness. Consider the modes to be the forward and backward waves on two identical sections of transmission line connected by a transducer for which reciprocity applies. Then P_a describes power flowing toward the transducer from the left, P_d describes power flowing away from the transducer at the right, Q_a describes power flowing away from the transducer at the left, and Q_d describes power flowing toward the transducer from the right. The lower signs apply in (3.15) and (3.16) and $(1+k^2)^{\frac{1}{2}}$ is greater than unity. Reference planes r_1 and r_3 coincide and reference planes r_2 and r_4 coincide.

Imagine the lines to have reflectionless terminations to the left and right. Reciprocity requires that the ratio of P_d to P_a when Q_d is zero be the same as the ratio of Q_a to Q_d when P_a is zero.

When $P_a=0$,

$$Q_a/Q_d=\frac{e^{-j\theta_3}}{(1+k^2)^{\frac{1}{2}}}. \tag{4.1}$$

When Q_d is zero,

$$Q_a=-\frac{k}{(1+k^2)^{\frac{1}{2}}}e^{j(\theta_1-\theta_2)}P_a, \tag{4.2}$$

so that

$$P_d/P_a=\frac{e^{j\theta_1}}{(1+k^2)^{\frac{1}{2}}}. \tag{4.3}$$

Thus, reciprocity requires that

$$\theta_3 = -\theta_1. \tag{4.4}$$

Accordingly, when reciprocity holds the equations for the most general transducer are

$$P_d = (1+k^2)^{\frac{1}{2}} e^{j\theta_1} P_a + k e^{j\theta_2} Q_a, \tag{4.5}$$

$$Q_d = (1+k^2)^{\frac{1}{2}} e^{-j\theta_1} Q_a + k e^{j\theta_3} P_a. \tag{4.6}$$

5. MOVING MEDIUM; NEGATIVE ENERGY

When a medium carrying a wave moves with respect to the observer, a wave may have a negative energy if it involves motion in the direction of motion of the medium. That is, the moving medium may have less energy when a wave travels in it than in the absence of a wave. This can be so because kinetic energy is proportional to the square of the total velocity and the medium can be rarefied in regions of highest velocity and compressed in regions of lowest velocity.

The direction of energy flow of the wave must be the direction of the group velocity. But, if the energy is negative the direction of power flow may be contrary to the group velocity. This is the case for the slower space-charge wave when the group velocity is in the direction of motion of the electrons. Chu,[4] and L. R. Walker in unpublished work, have pointed out the negative energy of the slower space-charge wave.

The negative energy of the slower space-charge wave accounts for many typical microwave tube phenomena. Because of it, the parameter $(1+k^2)^{\frac{1}{2}}$ is greater than unity and we can have waves whose amplitude increases or decreases exponentially with distance. Examples are:

In a traveling-wave tube[5] the slow space-charge wave interacts with an electromagnetic wave moving in the same direction and gives rise to a growing and a decreasing wave. In this case the coupling is really continuous and we can approximate this continuous coupling with a chain of very many very short transducers, with k given as so much per unit length. The stop band occurs when the argument of the cosine in (3.21) is near zero and the modes which are coupled are nearly in synchronism. This is markedly different from the case of coupling between the forward and backward waves of a transmission line, in which the two waves are never in synchronism and a stop band can occur only for discrete coupling and never with continuous coupling.

In a double-stream amplifier[5] a slow space-charge wave of one electron stream interacts with a fast space-charge wave of another electron stream and gives rise to a growing and a decreasing wave.

In a space-charge-wave amplifier,[6] the slow and fast space-charge waves of one stream are coupled periodically and produce an increasing and a decreasing wave. In accordance with the remarks made in section 3, the region of "gain" or exponential change of amplitude with distance will occur when the difference of phase of the two waves over one period of the coupling (accelerating plus decelerating region, for instance) is near to $2n\pi$.

6. MOVING MEDIUM; ENERGY SOURCE AT TRANSDUCER

We can explain many interesting phenomena involving moving media in terms of waves with negative energies. Consider, however, a purely torsional wave in a rod translated parallel to its axis. The velocities associated with the wave are at right angles to the velocity of translation and the kinetic energies add. An analysis shows that when a torsional wave on a fixed rod is coupled purely by couples about the axes to the slower torsional wave on a parallel rod moving axially with respect to the first, no gaining wave results.

This may seem contrary to results obtained with a mechanical device built by C. C. Cutler, C. F. Chapman, and W. E. Mathews.[2] In this device, two torsional systems were coupled by means of magnets. The torsional systems extended around the rims of two bicycle wheels, and when the wheels rotated in opposite directions at a critical speed oscillations in the systems increased in amplitude with time. Although the situation is somewhat different from that which we have considered, the result does indicate the growth of an ac disturbance in the absence of negative energy, and similar linear torsional systems in relative motion might well exhibit exponentially increasing waves.

This can be explained by assuming that longitudinal forces associated with the coupling act so that when two modes are present on media in relative motion a force is present acting in the direction of motion, which can slow down the moving system and put energy into the waves. Let us consider this possibility.

Suppose that the modes described by P and Q are modes for media in relative motion, both carrying power in the same direction. Assume that energy is not conserved at the transducer, but power is produced or absorbed. Since all amplitudes present can be expressed in terms of P_a and Q_a, the power produced must be a function of these. It must be a real function of the complex quantities P_a and Q_a, so let us assume the most general real function of products of P's and Q's, described in terms of real parameters α, β, γ, δ.

$$\begin{aligned} &P_d P_d^* + Q_d Q_d^* \\ &\quad = P_a P_a^* + Q_a Q_a^* + \alpha(P_a^* Q_a^* + P_a Q_a) \\ &\qquad + \beta(P_a Q_a^* + P_a^* Q_a) + \gamma(P_a P_a^*) + \delta(Q_a Q_a^*). \end{aligned} \tag{6.1}$$

[4] L. J. Chu, Paper presented at the Institute of Radio Engineer's Electron Devices Conference, University of New Hampshire, June, 1951.

[5] J. R. Pierce, *Traveling Wave Tubes* (D. Van Nostrand Company, Inc., New York, 1950).

[6] Ping King Tien and Lester M. Field, Proc. Inst. Radio Engrs. **40**, 688–695 (1952).

By using (1.1) and (1.2) we obtain

$$P_aP_a^*(AA^*+DD^*-1-\gamma)+Q_aQ_a^*(BB^*+CC^*-1-\delta)+\{[P_aQ_a^*(AB^*+DC^*-\beta)]+[P_aQ_a^*(AB^*+DC^*-\beta)]^*\}-(P_a^*Q_a^*+P_aQ_a)\alpha=0. \quad (6.2)$$

If Q_a is zero and P_a is not zero, all terms of (6.2) but the first vanish and hence the coefficient of $P_aP_a^*$ must be equal to zero. Similarly, the coefficient of $Q_aQ_a^*$ must be equal to zero. The third term of (6.2) depends on the difference of the phase angles of P_a and Q_a while the fourth term depends on the sum of these phase angles. By adjusting the difference without changing the sum, we can make the third term zero without changing the fourth term, so α must be zero. Since all terms but the third have been shown to be equal to zero, the coefficient of $P_aQ_a^*$ must also be zero. Hence we have the relations

$$AA^*+DD^*-1-\gamma=0, \quad (6.3)$$

$$BB^*+CC^*-1-\delta=0, \quad (6.4)$$

$$AB^*+DC^*-\beta=0, \quad (6.5)$$

$$\alpha=0. \quad (6.6)$$

Only three parameters are possible: β, γ, and δ.

Let us now consider a chain of transducers, to which (2.5) applies. We can simplify matters by choosing special reference planes so that A, D, C are made real and as a consequence B is real. In doing this we do not really lose generality; we merely transfer some phase shifts from the transducer to θ_p, θ_q and the relative phases of P_n and Q_n. We then have from (6.3)–(6.6),

$$\gamma=A^2+D^2-1, \quad (6.7)$$

$$\delta=B^2+C^2-1, \quad (6.8)$$

$$\beta=AB+DC. \quad (6.9)$$

Now we note that (2.5) can give rise to unattenuated waves only if

$$A=C, \quad (6.10)$$

$$AC-BD=A^2-BD=1. \quad (6.11)$$

If this is so,

$$m^2+2Am\cos\left(\frac{\theta_q-\theta_p}{2}\right)+1=0. \quad (6.12)$$

In this case, the waves are always unattenuated if $A<1$, but if $A>1$ the amplitude changes exponentially with distance if

$$A\cos\left(\frac{\theta_q-\theta_p}{2}\right)>1. \quad (6.13)$$

Thus, the behavior of the waves as $(\theta_p-\theta_q)$ is varied depends only on how big A is; A is the only remaining parameter.

It is interesting to note that if (6.10) and (6.11) hold

$$\gamma=B(B+D), \quad (6.14)$$

$$\delta=D(B+D), \quad (6.15)$$

$$\beta=(1+BD)^{\frac{1}{2}}(B+D), \quad (6.16)$$

$$A=(1+BD)^{\frac{1}{2}}.$$

We note that for small couplings, that is, if

$$|B|\ll|A|, \quad (6.17)$$

$$|D|\ll|A|. \quad (6.18)$$

Then

$$\delta\ll\beta, \quad (6.19)$$

$$\gamma\ll\beta. \quad (6.20)$$

If we examine (6.1), we see that this implies that the power produced directly through P_a and Q_a is larger than that which will be produced when P_a or Q_a is zero (presumably, power-produced by interactions involving P_b and Q_b).

It seems that there are only two alternatives—either (6.14)–(6.20) are true for all coupled dissipationless systems with nonnegative energy in relative motion, or some such systems have no unattenuated modes. The author guesses that the former may be correct.

7. COMMENTS

It will not have escaped the reader that the physical content of this material is rather small and the phenomena to which it is applied are somewhat varied. Many of the equations could as well be applied to the focusing of light rays by periodic lens systems or to the motion of electrons in periodic fields, thus giving an even broader and perhaps more dubious generality to the work. When we look carefully even at the particular instances we have considered, however, we may find our sense of propriety violated. Thus, consider a space-charge-wave amplifier in which the parameters of the electron beam are not constant over any appreciable distance. In this case the uncoupled modes do not really exist; the system is all transducers with no space between them. This does not invalidate the analysis but it does mean that P and Q are not really the amplitudes of waves which exist in the physical device, but are rather convenient variables in terms of which the excitation of the electron stream can be expressed.

It would perhaps be fruitless to hope to find one picture which would fit like a glove a large group of phenomena or devices in all their varied forms. What this paper does purport to do is to point out certain formal similarities which are to some extent physically appealing.

The author wishes to acknowledge the help of L. R. Walker, which led to a considerable simplification and shortening of this work.

Electron Diffraction Study of the Structures of Antimony Electrodeposited on Antimony Cleavage Face

LING YANG
Metals Research Laboratory, Carnegie Institute of Technology, Pittsburgh, Pennsylvania
(Received June 29, 1953)

Thin layers of antimony of average thickness about one hundred angstroms, electrodeposited at 90°C and 2 ma/cm² on the (111) cleavage face (referred to pseudo-cubic axes) of an antimony single crystal, yield abnormal electron diffraction patterns. Analysis of these patterns shows that the deposit crystals are arranged in three azimuthal equivalent orientations, with their $(11\bar{1})$ instead of the (111) lattice planes parallel to the (111) antimony substrate. This is confirmed by comparing the patterns obtained experimentally with those constructed from the Laue zones of reinforcement of the electron beam scattered from three nonparallel atom rows. This observation is discussed in connection with the various factors controlling the crystal orientation in electrodeposited metals.

DURING a series of investigations of the crystal growth of electrodeposited metals, single crystal faces of several metals having different crystal structures were used as substrates, and the structures of thin layers of various metals electrodeposited on them were studied by electron diffraction. Abnormal diffraction patterns were obtained in a number of cases, one of which, antimony on antimony (111) cleavage face (referred to pseudo-cubic axes), has been analyzed and is described here.

EXPERIMENTAL

Antimony single crystal was prepared by melting chemically pure antimony metal in a graphite crucible (inside diameter 1 cm) having a taper bottom in an electric furnace at 700°C and then lowering the crucible through a temperature gradient of 20°C per cm at a rate of 1 cm per ten minutes. Cleaving was done with a sharp blade after the direction of the cleavage face in the crystal had been located by breaking the tip of the crystal. Figure 1 is an electron diffraction pattern taken from the (111) cleavage face with a Finch-type camera having a distance of 25 cm between the specimen and the photographic plate. The presence of clear Kikuchi lines in the pattern indicates that the crystal possessed a high degree of lattice perfection and that at least a large part of the cleavage face was highly smooth.

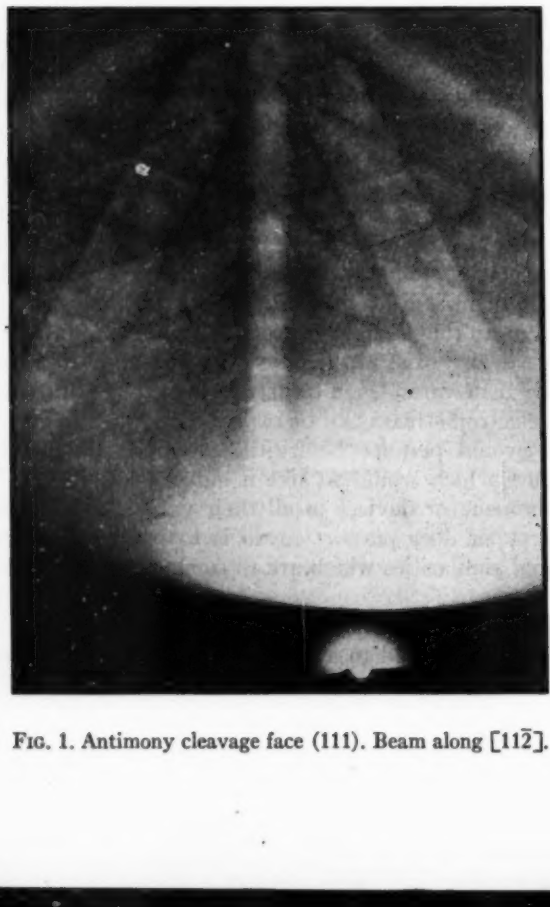


FIG. 1. Antimony cleavage face (111). Beam along $[11\bar{2}]$.

Antimony was electrodeposited on the freshly cleaved face from a bath containing 200 gm $SbCl_3$ and 300 cc concentrated HCl in one liter of the solution at 90°C and 2 ma/cm². The reagents used for preparing the solution were of the chemically pure grade, and the anode was electrolytic antimony. The bath was purified by electrolysis under the aforementioned conditions with a dummy antimony cathode before the deposition on the single crystal face. The average thickness of the deposit was estimated to be about 100A from the current passed, the time of deposition, and the surface area of the cathode, assuming 100 percent current efficiency. After the deposition the crystal was removed from the bath when the electric circuit was still on and washed immediately with 2 N HCl followed by distilled water.

When the deposit was examined in a Finch-type camera (specimen to photographic plate distance $L=47.5$ cm), the patterns shown in Figs. 2(a) and 2(b) were obtained. These patterns, however, are not those expected from a (111) arrangement of the deposited atoms on the (111) substrate. In order to know exactly the camera length $L\lambda$ (λ = the wavelength of the electron beam) when the patterns were taken and hence to calculate the spacing d of the lattice plane in the deposit crystals, which was parallel to the substrate surface, a little graphite powder was dusted on the surface of the specimen and a composite pattern Fig. 2(c) was taken. From the radius of the $11\bar{2}0$ ring of

graphite, the camera constant was evaluated, the basal axis length* of graphite being taken as 2.4613A.[1] Using this camera constant and the distances D between the central spot and the spots of successive order of diffraction N along the plane of incidence in Fig. 2(c), it was found, as shown in Table I, that the deposit crystals were oriented with their $(11\bar{1})$ instead of (111) lattice planes parallel to the (111) substrate surface.

Although Table I shows that the $(11\bar{1})$ planes of the deposit are parallel to the (111) of the substrate, there remain to be explained why in Figs. 2(a) and 2(b) some of the diffraction spots are doublets and why in Fig. 2(a) only the diffraction spots to the left of the plane of incidence are doublets. Since there is a number of different ways to put $Sb(11\bar{1})$ on Sb(111), it is therefore decided to calculate the patterns corresponding to the different types of arrangement of $Sb(11\bar{1})$ on Sb(111) in order to find out whether the abnormal characteristics

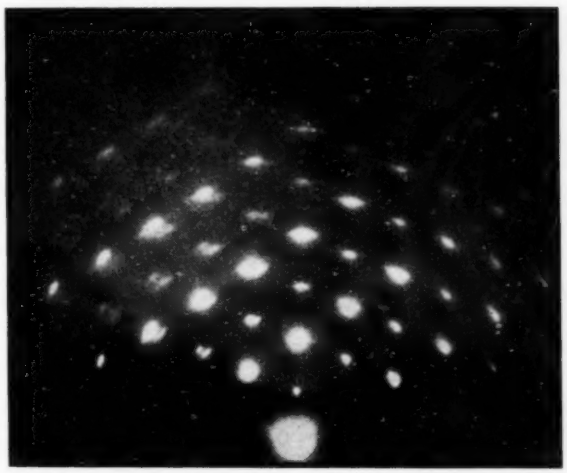


FIG. 2(a). Antimony electrodeposited on antimony (111). Average thickness about 100A. Beam along $[10\bar{1}]$.

of the diffraction patterns can be explained by the presence of some or all of these different types of arrangement.

CALCULATION OF THE PATTERNS CORRESPONDING TO THE DIFFERENT TYPES OF ARRANGEMENT OF Sb $(11\bar{1})$ ON Sb (111)

The antimony unit cell, referred to the pseudo-cubic axes, is shown in Fig. 3(a), and the atomic arrangements of the (111) and the $(11\bar{1})$ lattice planes are shown in Fig. 3(b). When the $(11\bar{1})$ is put on the (111) plane, it can easily be seen that the arrangement may be either like Fig. 4(a), in which the l_1 of $(11\bar{1})$ is parallel to the l_1 of (111) and there are three possible azimuthal equivalent orientations (Type A), or like Fig. 4(b), in

* All the lattice dimensions in this work are in absolute angstrom.

[1] W. Hume-Rothery, *The Structure of Metals and Alloys* (The Institute of Metals, London, 1945), p. 50.

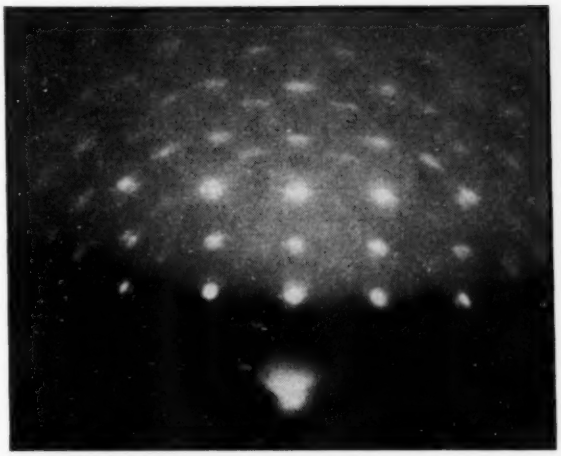


FIG. 2(b). Antimony electrodeposited on antimony (111). Average thickness about 100A. Beam along $[11\bar{2}]$.

which the l_2 of $(11\bar{1})$ is parallel to the l_1 of (111) and there are six possible azimuthal equivalent orientations (Type B). The deposit crystals may have either of these arrangements or both. In order to find out which

TABLE I. Determination of the lattice plane in the antimony deposit, which is parallel to the (111) antimony substrate [see Fig. 2(c)].

N	D(cms)	D/N	$d=(L\lambda)/(D/N)$ (A)	d(average)	$d_{Sb(111)}$* (calc.)
2	1.343	0.672	3.559		
3	2.022	0.674	3.549		
4	2.713	0.678	3.529		
5	3.392	0.678	3.529	3.542	3.541

Radius of $11\bar{2}0$ ring of graphite in Fig. 2(c) = 1.943 cms
Lattice spacing of graphite $(11\bar{2}0)$ = 1.231A
$L\lambda$ = camera constant = 1.943 × 1.231 = 2.392

* Calculated from the unit cell dimensions of Sb (see reference 1, p. 51), a = 4.5067A, α = 57°6′19″, when referred to rhombohedral structure, or a = 6.233A, α = 87°24′ when referred to pseudo-cubic structure.

of these arrangements occurs and whether such arrangement yields patterns similar to those found experimentally, the electron diffraction patterns for zero third Laue zone, corresponding to each of these two types of arrangement, have been worked out. This is

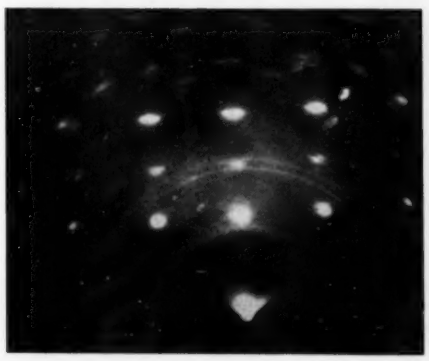


FIG. 2(c). Composite pattern of Fig. 2(b) and graphite.

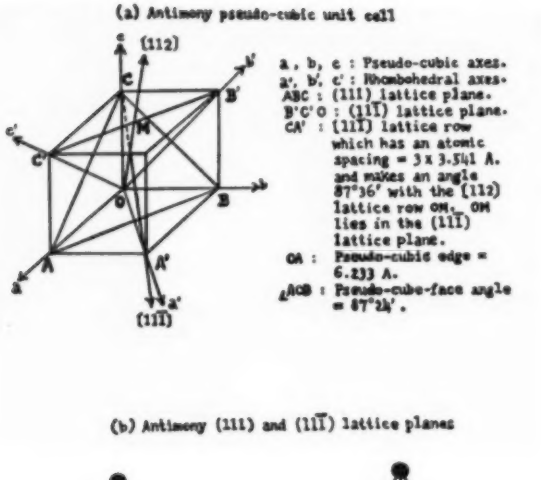

(a) Antimony pseudo-cubic unit cell
(112)
(111)
a, b, c : Pseudo-cubic axes.
a', b', c' : Rhombohedral axes.
ABC : (111) lattice plane.
B'C'O : (11$\bar{1}$) lattice plane.
CA' : [11$\bar{1}$] lattice row which has an atomic spacing = 3 x 3.541 A. and makes an angle 87°36' with the [112] lattice row OM. OM lies in the (11$\bar{1}$) lattice plane.
OA : Pseudo-cubic edge = 6.233 A.
∠ACB : Pseudo-cube-face angle = 87°24'.
(11$\bar{1}$)
(b) Antimony (111) and (11$\bar{1}$) lattice planes
60°
87°6'
l_1
l_2
l_3
l_4
Sb (111)
Sb (11$\bar{1}$)
l_1= 4.308 A.
l_2= 4.506 A.
= Sb atom.
l_3(bisecting l_1) = 3.961 A.
l_4(bisecting l_2) = 3.789 A.

FIG. 3. Antimony pseudo-cubic unit cell and its (111) and (11$\bar{1}$) lattice planes.

done by considering the Laue zones of reinforcement of the electron beam scattered from three nonparallel atom rows and the structural factor of each reinforced diffraction. The atom rows chosen for this purpose are: one nearly at a right angle to the specimen surface, one in the surface of the specimen and nearly perpendicular to the direction of the electron beam, and one nearly parallel to the direction of the electron beam. The data used for the calculation are listed in Table II and the calculated patterns are shown in Figs. 5(a) and 5(b) for Type A arrangement and in Figs. 6(a) and 6(b) for Type B arrangement.

When these calculated patterns are compared with the experimentally obtained patterns Figs. 2(a) and 2(b), it can be seen that both Fig. 5(a) and Fig. 6(a) agree completely with Fig. 2(a), after the fact is taken into account that in Fig. 5(a) and Fig. 6(a) each pair of adjacent spots at the right of the plane of incidence becomes merged into a single arc. However, when Figs. 2(b), 5(b), and 6(b) are compared, it can be seen that Fig. 5(b) but not Fig. 6(b) resembles Fig. 2(b). Although in Fig. 6(b) it is possible for the two upper spots of each cluster of spots to join into a single arc, yet the arc formed should extend over the top of the lower spot. This is obviously not so in Fig. 2(b). This indicates that the deposit crystals choose the Type A arrangement in which the (11$\bar{1}$) planes of the deposit crystals are arranged in three azimuthal equivalent orientations with respect to the (111) plane of the substrate and the l_1 of (11$\bar{1}$) is parallel to the l_1 of (111). The above analysis therefore not only accounts sucessfully for the abnormal characteristics in the electron diffraction patterns obtained experimentally but also

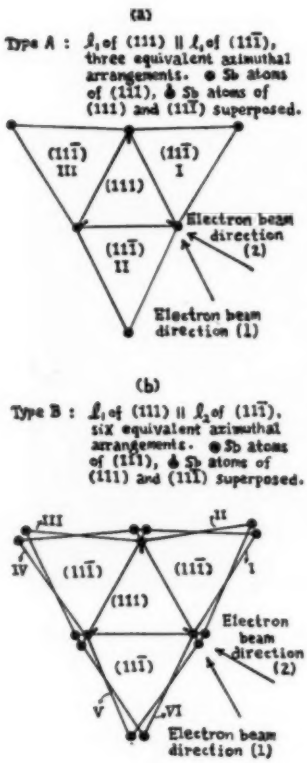

(a)
Type A : l_1 of (111) ∥ l_1 of (11$\bar{1}$), three equivalent azimuthal arrangements. ● Sb atoms of (11$\bar{1}$), ◆ Sb atoms of (111) and (11$\bar{1}$) superposed.
(11$\bar{1}$) III
(11$\bar{1}$) I
(111)
(11$\bar{1}$) II
Electron beam direction (2)
Electron beam direction (1)
(b)
Type B : l_1 of (111) ∥ l_1 of (11$\bar{1}$), six equivalent azimuthal arrangements. ● Sb atoms of (11$\bar{1}$), ◆ Sb atoms of (111) and (11$\bar{1}$) superposed.
III
II
IV
(11$\bar{1}$)
(11$\bar{1}$)
I
(111)
Electron beam direction (2)
(11$\bar{1}$)
V
VI
Electron beam direction (1)

FIG. 4. Possible arrangements of antimony (11$\bar{1}$) on antimony (111).

TABLE II. Data used for the calculation of the patterns corresponding to the different types of arrangement of Sb(11$\bar{1}$) on Sb(111) [see Figs. 4(a) and 4(b)].

Beam direction	Type of arrangement	Lattice rows in deposit surface[a] Nearly ⊥ to beam (L_1) Spacing	α	Nearly ∥ to beam (L_2) Spacing	β	Lattice rows nearly ⊥ to deposit surface[a] Always (11$\bar{1}$) Spacing	ρ
	A(I)	$4h_1$	0	$2l_1$	0	$3l_3$	−2°24′
(1)	A(II)	$4h_2$	0	$2l_2$	+1°27′	$3l_3$	+1°12′
	A(III)	$4h_2$	0	$2l_2$	−1°27′	$3l_3$	+1°12′
	A(I)	$2l_1$	0	$4h_1$	0	$3l_3$	0
(2)	A(II)	$2l_2$	0	$4h_2$	−1°29′	$3l_3$	−2°4′
	A(III)	$2l_2$	0	$4h_2$	+1°29′	$3l_3$	+2°4′
	B(I)	$4h_2$	0	$2l_2$	0	$3l_3$	+1°9′
	B(II)	$4h_2$	0	$2l_2$	0	$3l_3$	+1°9′
(1)	B(III)	$4h_1$	0	$2l_1$	−1°27′	$3l_3$	−2°24′
	B(IV)	$4h_2$	0	$2l_2$	+2°54′	$3l_3$	+1°15′
	B(V)	$4h_2$	0	$2l_2$	−2°54′	$3l_3$	+1°15′
	B(VI)	$4h_1$	0	$2l_1$	+1°27′	$3l_3$	−2°24′
	B(I)	$2l_2$	0	$4h_2$	−1°25′	$3l_3$	+2°3′
	B(II)	$2l_1$	0	$4h_1$	+1°27′	$3l_3$	0
(2)	B(III)	$2l_2$	0	$4h_2$	−1°29′	$3l_3$	−2°6′
	B(IV)	$2l_2$	0	$4h_2$	+1°29′	$3l_3$	+2°6′
	B(V)	$2l_1$	0	$4h_1$	−1°27′	$3l_3$	0
	B(VI)	$2l_2$	0	$4h_2$	+1°25′	$3l_3$	−2°3′

[a] α = angle between the projection of L_1 on photographic plate and the shadow edge of the specimen. β = angle between L_2 and the beam direction. The positive or negative sign means L_2 is at the right or the left of the beam. ρ = angle between the projection of [11$\bar{1}$] on the photographic plate and the plane of incidence. The positive or negative sign means the projection is at the right or the left of the plane of incidence. h_1 = 3.961A, h_2 = 3.789A, l_1 = 4.308A, l_2 = 4.506A, l_3 = 3.541A.

helps to differentiate the two possible types of arrangement of the deposit crystals.

DISCUSSION

Previous investigations[2] on the crystal growth of electrodeposited metals have shown that the structures of the initial thin deposit are chiefly determined by those of the substrate and that the structures of the thicker deposit are controlled by the bath conditions. It has also been known[2] that in thick (30 000A) polycrystalline antimony electrodeposited under the bath conditions used in this work on the mechanically polished surface of brass, which is void of any substrate influence, the deposit crystals are preferentially oriented with their $(11\bar{1})$ planes [or (100) planes when referred to rhombohedral axes] parallel to the substrate surface. This indicates that the $(11\bar{1})$ orientation is the one favored by the prevailing conditions of deposition. As shown in this work, the substrate influence which favors (111) arrangement of the deposited atoms seems to be outweighed by the influence of the bath conditions even at the very beginning of the deposition. This may be partly because of the weak surface force on the cleavage face and partly because of the similarity in the atomic arrangements on the (111) and the $(11\bar{1})$ lattice planes [see Fig. 3(b)]. The deposited atoms, by taking the $(11\bar{1})$ orientation, can satisfy the requirement of the bath conditions and at the same time maintain an arrangement closely resembling that favored by the substrate surface. Orientation similar to that found in this work has also been observed in antimony electrodeposited under the above-mentioned bath conditions on bismuth (111) cleavage face. On bismuth (100) plane (referred to pseudo-cubic axes), however, the initial layers of antimony deposit take the normal (100) orientation, probably because the (100) plane has stronger surface force than the (111) cleavage face.

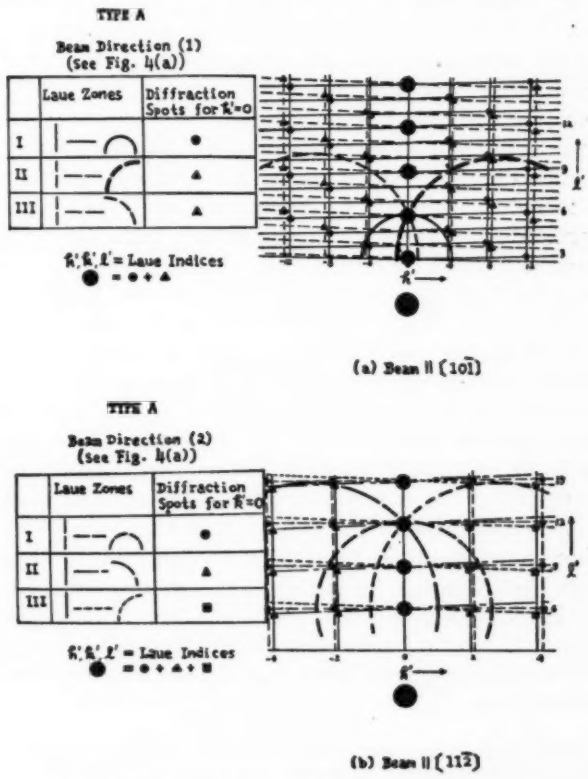


FIG. 5. Calculated patterns of antimony $(11\bar{1})$ on antimony (111), Type A.

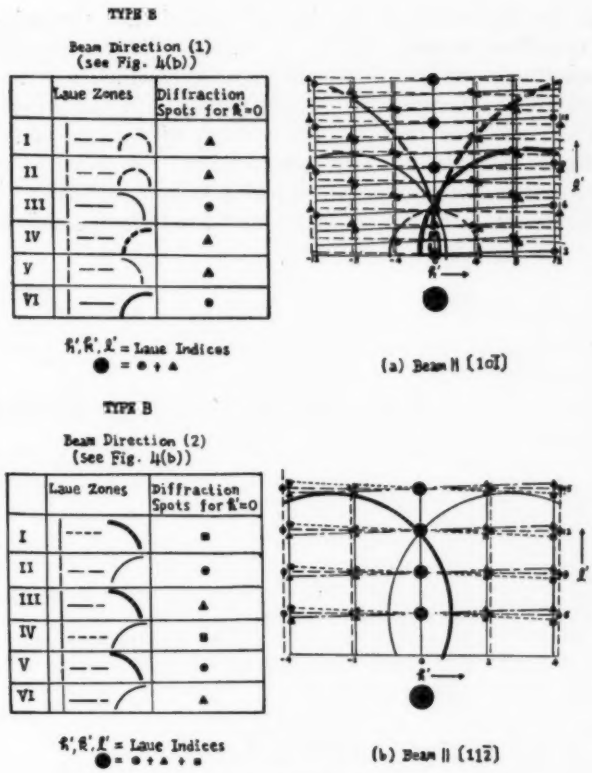


FIG. 6. Calculated patterns of antimony $(11\bar{1})$ on antimony (111), Type B.

Another example illustrating the competition between the substrate influence and the influence of the bath conditions in controlling the orientation of electrodeposited metal crystals can be found in the electrodeposition of nickel on copper single crystal faces. Previous study[2] by electron diffraction showed that when nickel was electrodeposited on copper (100) face under bath conditions favoring the development (100) orientation, the deposit persisted monocrystalline to a thickness of several microns. Recently, Leidheiser and Gwathmey[3] showed by means of x-rays that under such bath conditions nickel deposited on copper (100) face was monocrystalline even at a thickness of 0.01 inch. On copper (111) face, however, these authors found that the deposit became polycrystalline almost

[2] Finch, Wilman, and Yang, Trans. Faraday Soc. 43A, 144 (1947).

[3] H. Leidheiser, Jr., and A. T. Gwathmey, J. Electrochem. Soc., 98, 225 (1951).

at the beginning of the deposition and that the polycrystalline deposit rapidly assumed a (100) preferred orientation as the thickness increased, which was the orientation favored by the prevailing bath conditions. This indicates that for the growth of thicker metal single crystals by electrodeposition, the chance of success is better if considerations have been given to the selection of the substrate and the bath conditions so that both of them favor the development of the same type of orientation in the deposit.

ACKNOWLEDGMENT

I wish to thank Dr. H. Wilman of the Applied Physical Chemistry Laboratory, Imperial College of Science and Technology, London, England, for many stimulating discussions.

JOURNAL OF APPLIED PHYSICS VOLUME 25, NUMBER 2 FEBRUARY, 1954

Turbulent Processes as Observed in Boundary Layer and Pipe*

G. B. SCHUBAUER
National Bureau of Standards, Washington, D. C.
(Received June 17, 1953)

Certain statistical properties of turbulence observed in a boundary layer and in fully developed pipe flow are compared. Differences and likenesses are shown, and in particular certain common properties are found which appear to be indicative of processes governing the finer structure and controlling the origin, form, and life cycle of turbulent motions. An attempt is made to point out some of these processes. Attention is called to a region of high turbulent activity near the wall.

SYMBOLS

x = distance along wall in direction of mean flow.
y = distance normal to wall, measured from wall.
z = coordinate normal to x and y.
U, V = x and y components of mean velocity.
U_1 = velocity of free stream or velocity at center of pipe or channel.
u,v,w = instantaneous turbulent velocities at a point in the x, y, and z directions, respectively.
$\langle u^2\rangle_{Av}, \langle v^2\rangle_{Av}, \langle w^2\rangle_{Av}$ = temporal mean-square turbulent velocities at a point.
δ = thickness of boundary layer; also radius of pipe or half-width of channel.
ν = kinematic viscosity.
ρ = density.
τ_0 = shearing stress at wall.
$U_\tau = (\tau_0/\rho)^{\frac{1}{2}}$ = friction velocity.
$R = \delta U_1/\nu$ = Reynolds number.
$y^* = yU_\tau/\nu$ = wall-distance parameter.
t = time.
n = frequency, cycles per sec.
k_1 = one-dimensional wave number in x direction (k represents three-dimensional wave number).
$F_u(k_1), F_v(k_1), F_w(k_1)$ = one-dimensional energy spectral functions for the u, v, and w components, respectively. For example $\langle u^2\rangle_{Av} = \int_0^\infty F_u(k_1)dk_1$.
$F(k)$ = three-dimensional energy spectral function.

1. INTRODUCTION

THE purpose of the present paper is to show how measurements of turbulence made in two distinctly different flow fields combine to reveal something of the behavior and of the life cycle of turbulent motions. The data are taken from two investigations[1,2] recently conducted at the National Bureau of Standards under the sponsorship of the National Advisory Committee for Aeronautics. One concerned itself with a turbulent boundary layer along a flat plate with uniform pressure, and the other dealt with fully developed turbulent flow in a straight round pipe.[1,2] It suffices to mention here that hot-wire techniques were employed in both cases, and for the benefit of hot-wire instrumentation the flow fields were made as large as practicable and speeds were held to moderate values. The boundary layer was 3 inches thick; the radius of the pipe was 4.86 inches; the speed of the free stream was 50 ft per sec; and the speeds at the center of the pipe were 10 to 100 ft per sec. The measurements made were not identical in the two cases, but in general consisted of mean-square values of the velocity fluctuations and their derivatives, double and triple products of fluctuation components, and spectra. Particular attention was paid to quantities entering into the energy balance. Only certain quantities are selected for comparison here.

* This investigation was sponsored by the National Advisory Committee for Aeronautics. Paper presented at the Symposium of the Division of Fluid Dynamics, American Physical Society, January 24, 1953.

[1] P. S. Klebanoff, "Characteristics of turbulence in a boundary layer with zero pressure gradient," Natl. Bur. Standards (U. S.) Report 2454 (April, 1953) (submitted to the National Advisory Committee for Aeronautics for publication).
[2] John Laufer, "The structure of turbulence in fully-developed pipe flow," Natl. Advisory Comm. Aeronaut. Tech. Notes 2954 (June, 1953).

It is often difficult to assess the accuracy of hot-wire measurements. In these two cases it is believed that many of the errors that so often plague such measurements were avoided or at least reduced, by a fortunate choice of conditions and the care exercised by the experimenters. Accurate and sufficient data are essential if interpretations are to have a lasting quality. Perhaps only future results will tell whether or not these conditions were satisfied. Deficiencies are apparent, and it is hoped that, whereas permanence may be lacking, the conclusions will have a thought-provoking value.

2. GENERAL FORM OF FLOW FIELDS

It is generally known that the mean velocity distribution in a boundary layer is different from that in pipes and channels. The difference is most distinct at the higher Reynolds numbers where it is not so likely to be obscured by Reynolds-number effects. Figure 1 shows this characteristic difference. It is seen that the profiles for the channel[3] and the pipe are essentially the same in spite of the difference in geometry, but that the profile for the boundary layer stands apart by far more than experimental scatter. The former obey a logarithmic law with a fair degree of approximation, whereas the boundary-layer profile shows marked deviations from a log law for $y/\delta > 0.2$.

The cause of this singular behavior of the boundary layer is now attributed to the very irregular outer limit of its boundary, as shown schematically in Fig. 2. Recent observations[1,4–7] have established the validity

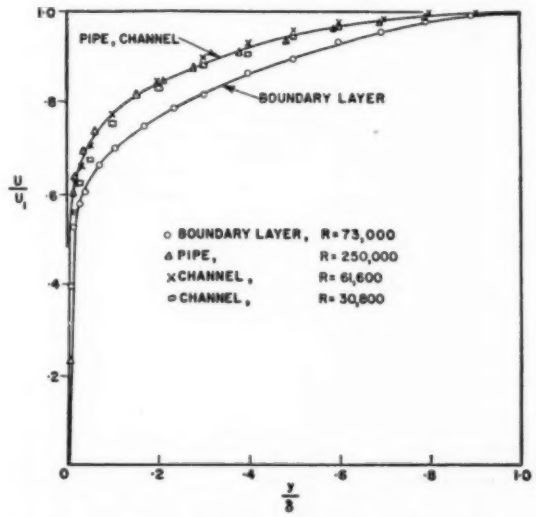


FIG. 1. Mean-velocity profiles for boundary layer, pipe, and channel.

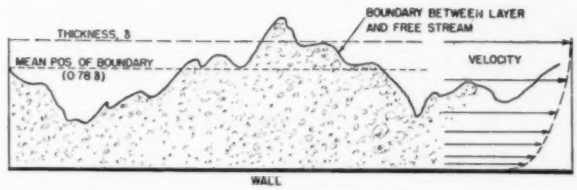


FIG. 2. Schematic diagram of boundary layer.

of such a picture whenever a turbulent flow is bounded by a free stream. A hot-wire anemometer placed in the neighborhood of the mean position of the edge would sense intermittently turbulent and nonturbulent flows as the irregular outline moved past with approximately the speed of the free stream. The fraction of the time that the flow is found to be turbulent defines an intermittency factor, commonly denoted by γ. This factor has been accurately determined, and its distribution is found to be closely represented by a Gaussian-integral curve centered at 0.78δ, δ being specified in the usual way as the distance from the wall where the mean velocity has reached the free-stream velocity. It is found that turbulent flow occasionally goes beyond δ and exceeds this limit 6 percent of the time. On the average the edge occupies the position 0.78δ, but from instant to instant it is statistically distributed about this line with close approach to a Gaussian distribution. The nonturbulent fluid rarely penetrates deeper than 0.4δ, and the turbulent outcroppings rarely go beyond 1.2δ. The region beyond 0.4δ is defined as the intermittent region.

The sharp division is the instantaneous limit of the turbulent motions. The boundary-layer thickness specifies the region over which a mean shear exists. It is, therefore, a measure of the scale of the shear flow even though it includes regions of nonturbulent flow, and in this sense is a dimension similar to the radius of a pipe or the half width of a channel. The difference lies in the inclusion of turbulent and nonturbulent portions within this shear region in the boundary layer but completely turbulent flow only, in the case of the pipe and channel.

The region beyond the sharp limit of the turbulent flow is assumed to be free stream. Oscillograms of velocity fluctuations clearly show the turbulent and nonturbulent regions with, however, some disturbances in the nonturbulent regions. Close inspection of the records show a step in the velocity which is taken as evidence that the mean velocity is lower in the turbulent regions.

Information regarding these phenomena is by no means complete. It is of interest, therefore, to compare in Fig. 3 the turbulent energy distribution in the pipe with that in the boundary layer. For the boundary

[3] John Laufer, "Investigation of turbulent flow in a two-dimensional channel," Natl. Advisory Comm. Aeronaut. Tech. Notes 2123 (July, 1950).

[4] Stanley Corrsin, "Investigation of flow in an axially symmetrical heated jet of air," National Advisory Committee for Aeronautics Wartime Report W–93 (1943).

[5] A. A. Townsend, Australian J. Sci. Research. Ser. A.2, No. 4, 451–468 (1949).

[6] A. A. Townsend, Proc. Cambridge Phil. Soc. **47**, Pt. 2, 375–395 (1950).

[7] Stanley Corrsin and Alan L. Kistler, "The free stream boundaries of turbulent flows," JHU Report, Natl. Advisory Comm. Aeronaut. Contract NAw–6093 (January, 1953).

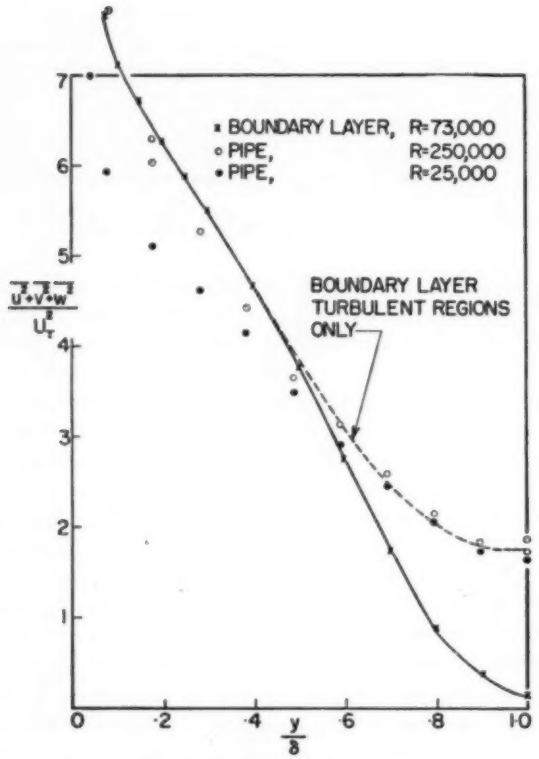


FIG. 3. Distribution of turbulent energy across boundary layer and pipe.

layer the values represent a mean including both turbulent and nonturbulent regions. If we now assume that no fluctuations exist in the nonturbulent regions and ignore a possible effect of a velocity step, division by γ yields the turbulent energy in the turbulent regions only. This is shown by the broken curve in Fig. 3, which is seen to coincide closely with the distribution in the pipe. A similar behavior is shown for each of the components in Fig. 4. The only deviation from remarkable agreement in these two figures occurs in the low Reynolds-number pipe values, and this is believed to be purely a Reynolds-number effect.

It would appear that such close agreement could not occur unless the average turbulent energy was distributed through the turbulent regions of a boundary layer just as it is in the fully turbulent section of a pipe. The alternative, namely, that the distributions are not the same and that fluctuations in the nonturbulent regions and velocity steps just happen to bring about agreement, would seem to be remote. Here, then, is an indication that disturbances in the free stream and velocity steps, even though observed on records, do not make a measurable contribution to the fluctuation energy. What is of deeper significance, perhaps, is the close similarity exhibited by the turbulent structure of these two flow fields. Evidently the average turbulent energy does not depend on position relative to the sharp boundary but on position within the shear layer δ. It is worthy of note that the turbulent energy does not vanish toward the sharp edge but maintains an average value like that in a pipe at the same relative distance from the wall.

3. CONDITIONS NEAR WALL

The most important outcome of the investigations reported in references 1 and 2 is the revelation that a region in the neighborhood of the hypothetical laminar sublayer is one of high activity for turbulence. It is found that not only does a high rate of dissipation occur here but a high rate of turbulence production exists as well.

Before discussing the figures, it is pointed out that the appropriate scales of velocity and length near the wall are taken to be U_τ and ν/U_τ respectively. All quantities when rendered dimensionless by means of these scales are then assumed to be functions of the customary wall-distance parameter, denoted by y^* and defined by

$$y^* = yU_\tau/\nu.$$

Such dimensionless quantities appear in Figs. 5 to 10. Comparisons between pipe and boundary-layer results are shown in these figures, with the exception of Fig. 7, where slight differences believed to be experimental are concealed by an average curve. The plotted points represent values derived from observations but are not an indication of the number of observations nor of the exact position at which a direct observation was made. The region in which many important events occur ex-

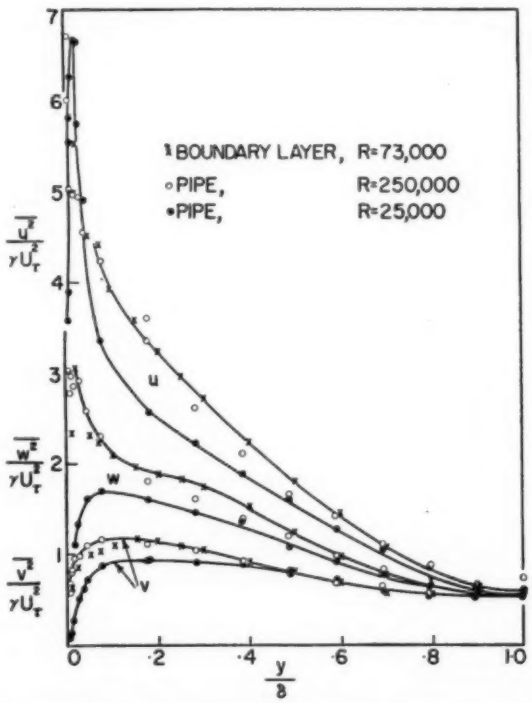


FIG. 4. Distribution of turbulent energy in components, pipe, and turbulent regions of boundary layer compared. γ is intermittency factor for boundary layer.

tends from the wall to distance y^* in the neighborhood of 100. This represents a small part of the shear layer, amounting to about 1 percent of δ at $R=250\,000$ and 10 percent at $R=25\,000$.

Chief interest here centers around energy relations. Figure 5 shows the direct dissipation of mean-flow energy compared to the turbulence production. Here are plotted the rates per unit mass given in the dimensionless forms:

$$\text{Direct dissipation}=\frac{\nu^2}{U_\tau^4}\left(\frac{dU}{dy}\right)^2;$$

$$\text{Turbulence production}=\frac{\nu}{U_\tau^4}\langle uv\rangle_{Av}\frac{dU}{dy}.$$

The process of dissipation of energy to more degenerate forms is usually pictured as a transfer of the energy from mean flow to turbulence and finally to heat. Here, however, we see that 40 percent of the mean-flow is dissipated directly as heat without going through the stage of turbulence. Only 60 percent of the energy lost by the mean flow produces turbulence. Of this amount, one-half is produced in the region of direct dissipation, which extends up to $y^*\approx 60$. The remaining one-half is distributed at a low, and gradually decreasing, rate throughout the remainder of the layer. It is worth noting that the maximum rate of production takes place at what is normally considered the edge of the laminar sublayer, and at this point the direct dissipation and turbulence production go on at the same rate.

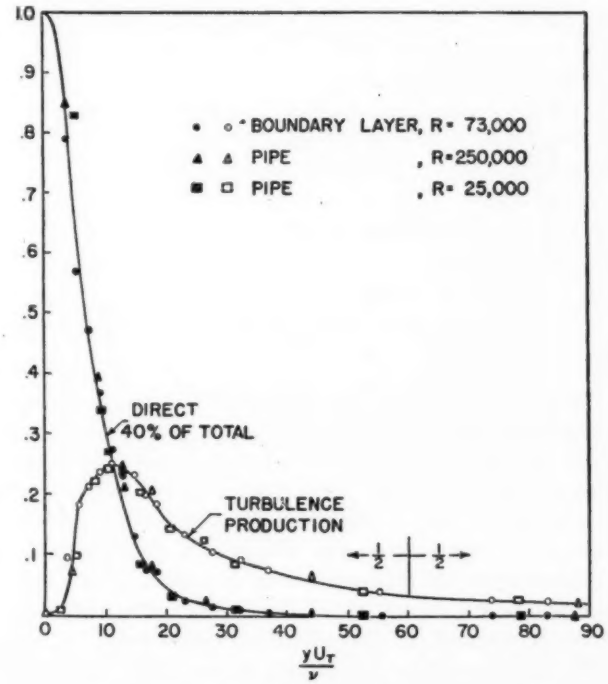


FIG. 5. Rate of loss of mean-flow energy, boundary layer, and pipe compared.

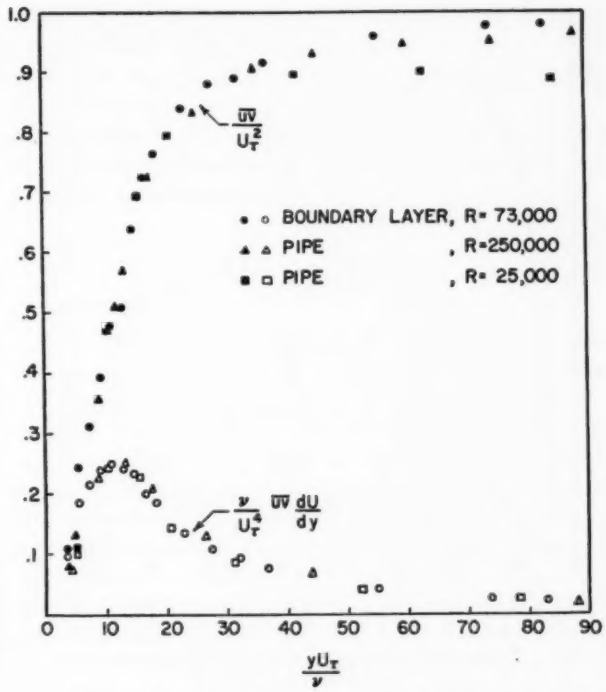


FIG. 6. Turbulent shearing stress and rate of turbulence production compared.

Figure 6 shows another illuminating phenomenon. As we leave the wall, production of turbulent energy reaches its maximum where the Reynolds stress is still very low, and the stress continues to increase while production is falling. This is a reflection of the tendency for the dissipation region to be crowded toward the wall by the less impeded turbulent mixing farther out. Where turbulent mixing can take place freely, the layers tend to be bound together, and the tendency is to approach a nonslipping, nondissipating mean flow. Not until the wall is closely approached is the mixing sufficiently impeded by the wall to allow the grip between laminae to be relaxed and a high rate of shear to set in. Here we find the so-called laminar sublayer with its high rate of shear and dissipation. We see now that part of this dissipative process is a high rate of turbulence production.

This brings us to the important question, how are turbulent motions produced to maintain the turbulent state and in what form do they first appear. The energy equations written separately for the three components u, v, and w tell us that the u component gets energy directly from the mean flow through the term

$$\langle uv\rangle_{Av}\frac{\partial U}{\partial y}.$$

The equations for v and w components lack this term. The three equations contain, respectively, the terms

$$\left\langle u\frac{\partial p}{\partial x}\right\rangle_{Av};\quad \left\langle v\frac{\partial p}{\partial y}\right\rangle_{Av};\quad \left\langle w\frac{\partial p}{\partial z}\right\rangle_{Av},$$

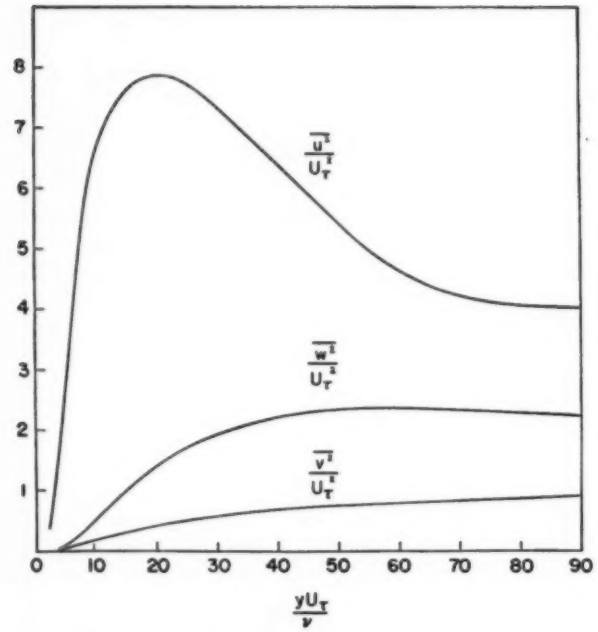


FIG. 7. Relative turbulent energies in three components near wall.

which express the rate of exchange of energy from one component to another. It is evident then that the lateral components v and w get energy only by exchange. The flow of energy is, therefore, from mean flow to u motions and from u motions to lateral motions.

In the interest of understanding the process it is helpful to have a physical picture. Experiment helps us indirectly with a picture by giving us statistical behaviors to which a picture must conform. For the details we can proceed on intuition and the knowledge that turbulent motions must be in the form of continuous, connected, intermingling currents that constantly interfere with one another and with solid walls and so change speed and direction in response to local, instantaneous pressure fields. The currents have every conceivable shape, some wandering far with little turning, others wrapping about themselves as a small eddy. Such a squirming assembly of currents large and small, interfering with themselves and with solid boundaries is about the only realistic picture of turbulence that can be drawn. Now let us ask what happens when this assembly is given a mean rate of shear. At once we see that currents traveling in the direction of the gradient (y direction) will find themselves taking on new relative motions with respect to surroundings, these being in the form of added components of motion parallel to the mean flow. In short, all movements in the y direction involve the setting up of new motions in the x direction. Turbulent motions are, therefore, born in the form of u motions.

The newly born turbulence is in reality only bits of mean flow out of the positions that formerly constituted a smooth gradient. Some pre-existing v motions were the displacing agents. The u motions agitate the surrounding fluid to produce new lateral motions. In other words they create local pressure fields, setting up the pressure gradients that constitute the transfer terms which transmit part of their energy to v and w motions. The v motions again displace the mean-gradient flow, and the process continues.

We have already seen in Fig. 4 that most of the energy remains in the u component. This is shown even more strikingly for the region under examination in Fig. 7. This figure also shows that in the neighborhood of $y^*=12$, where production is a maximum, the energy in the v component is small. Evidently production rests more on a large supply of energy made available by a mean-velocity gradient than upon agitation.

Another aspect of the life cycle of turbulence is shown in Fig. 8. Here we see that the region of maximum production is also the region of maximum dissipation. In fact, throughout the very narrow region in which 50 percent of the turbulence is produced, the dissipation very nearly balances the production. Apparently much of the turbulence dies in infancy. The conclusions that can be drawn in regard to the life span of turbulence and its movement from point of origin depend greatly on what we know about the various energy terms entering into the energy balance.

4. MEASUREMENT OF DISSIPATION AND ENERGY BALANCE

The equation of energy for three-dimensional turbulent motions in two-dimensional, incompressible mean flow is

$$\underset{\text{I}}{\langle uv\rangle_{Av}\frac{\partial U}{\partial y}} \underset{\text{II}}{-\nu(\langle u\nabla^2 u\rangle_{Av}+\langle v\nabla^2 v\rangle_{Av}+\langle w\nabla^2 w\rangle_{Av})} \underset{\text{III}}{+\frac{1}{2}\frac{\partial}{\partial y}(\langle vu^2\rangle_{Av}+\langle v^3\rangle_{Av}+\langle vw^2\rangle_{Av})} \underset{\text{IV}}{+\frac{1}{\rho}\frac{\partial\langle pv\rangle_{Av}}{\partial y}} + \frac{1}{2}\Big[\underset{\text{V}}{U\frac{\partial}{\partial x}(\langle u^2\rangle_{Av}+\langle v^2\rangle_{Av}+\langle w^2\rangle_{Av}) + V\frac{\partial}{\partial y}(\langle u^2\rangle_{Av}+\langle v^2\rangle_{Av}+\langle w^2\rangle_{Av})}\Big]=0, \quad (1)$$

where $\nabla^2=\partial^2/\partial x^2+\partial^2/\partial y^2+\partial^2/\partial z^2$. The equivalent form of the equation for axial symmetry is given in reference 2. Each term represents the rate of change of turbulent energy per unit mass at a given point. They are: I. Production. II. Dissipation. III. Kinetic energy diffusion. IV. Pressure energy diffusion. V. Convection.

Aside from the pressure energy diffusion, which cannot yet be measured, all of the above quantities have been determined by hot-wire techniques with various degrees of accuracy. Production is the easiest to determine and is subject to least error especially in pipe flow. Dissipation is the most difficult to determine but

reasonably successful attempts to measure it will be discussed in some detail.

Actually term II in Eq. (1) contains quantities that cannot be measured. It is instructive to note that the form of this term conceals the mechanism by which heat is evolved by internal friction. While it is customarily called the dissipation term in relation to the other terms of Eq. (1), it is in reality an expression of the work done by viscous forces on the boundaries of a volume of unit mass to change its kinetic energy as a whole. By transforming into an equivalent form the heat producing mechanism is brought into view, and terms appear that can be measured. The equivalent form is written

$$\tfrac{1}{2}\nu\frac{\partial^2}{\partial y^2}(\langle u^2\rangle_{Av}+\langle v^2\rangle_{Av}+\langle w^2\rangle_{Av})$$

$$-\nu\Big[\Big\langle\Big(\frac{\partial u}{\partial x}\Big)^2\Big\rangle_{Av}+\Big\langle\Big(\frac{\partial v}{\partial x}\Big)^2\Big\rangle_{Av}+\Big\langle\Big(\frac{\partial w}{\partial x}\Big)^2\Big\rangle_{Av}$$

$$+\Big\langle\Big(\frac{\partial u}{\partial y}\Big)^2\Big\rangle_{Av}+\Big\langle\Big(\frac{\partial v}{\partial y}\Big)^2\Big\rangle_{Av}+\Big\langle\Big(\frac{\partial w}{\partial y}\Big)^2\Big\rangle_{Av}$$

$$+\Big\langle\Big(\frac{\partial u}{\partial z}\Big)^2\Big\rangle_{Av}+\Big\langle\Big(\frac{\partial v}{\partial z}\Big)^2\Big\rangle_{Av}+\Big\langle\Big(\frac{\partial w}{\partial z}\Big)^2\Big\rangle_{Av}\Big]. \quad (2)$$

Omitted in this expression are two insignificant terms of the form of the first involving second derivatives with respect to x and z.

The first term of expression (2), when summed across the shear region, is zero. It therefore represents a migration of energy from one region to another. The second term is the form taken by the expression for the dissipation of turbulent energy when the turbulence is homogeneous. Actually the effect of inhomogeneity is so small that our second term is practically an expression of the rate of consumption of turbulent energy at each point by viscous action. This, however, is incidental since our task is to evaluate term II in Eq. (1). The important fact is that expression (2) involves terms that can be measured.

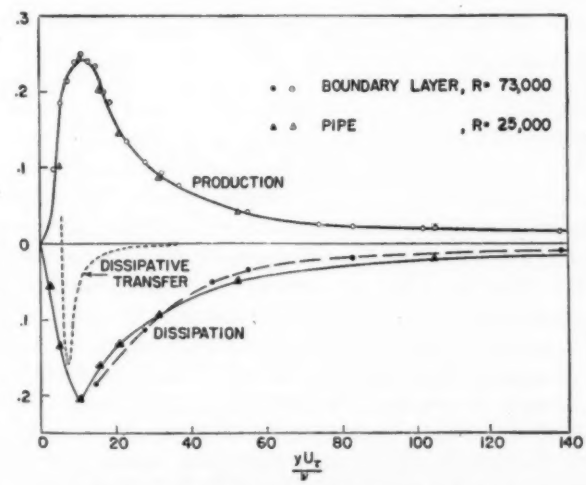


FIG. 8. Production and dissipation of turbulent energy compared.

In isotropic and homogeneous turbulence the task of obtaining the dissipation is greatly simplified. Here expression (2) is reduced to

$$15\nu\Big\langle\Big(\frac{\partial u}{\partial x}\Big)^2\Big\rangle. \quad (3)$$

In line with general thinking it was at first assumed that dissipation would be restricted to the smaller-scale motions among which isotropy and homogeneity possibly exist, and expression (3) was used. The turbulent dissipation found in this way when summed over the cross section of the pipe was less than the sum of the production by a factor 2.5. A slightly smaller error was found for the boundary layers when due account was taken of the convection term. In both cases indications were that the results were grossly in error near the wall. Accordingly, an effort was made to measure as many terms in expression (2) as possible. The terms

$$\Big\langle\Big(\frac{\partial u}{\partial x}\Big)^2\Big\rangle_{Av},\ \Big\langle\Big(\frac{\partial v}{\partial x}\Big)^2\Big\rangle_{Av},\ \Big\langle\Big(\frac{\partial w}{\partial x}\Big)^2\Big\rangle_{Av}$$

presented no unusual difficulties, for they involved the established practice of measuring time derivatives and using the conversion

$$\partial u/\partial t = U\partial u/\partial x.$$

The other mean-square derivatives involved the measurement of space correlations, and only u-sensitive probes could reach the highly important region near the wall. Consequently of the remainder only

$$\Big\langle\Big(\frac{\partial u}{\partial y}\Big)^2\Big\rangle_{Av} \text{ and } \Big\langle\Big(\frac{\partial u}{\partial z}\Big)^2\Big\rangle_{Av}$$

could be measured. For want of a better procedure it was assumed that mean-square derivatives with respect to a given coordinate separately satisfied the isotropic relations

$$2\Big\langle\Big(\frac{\partial v}{\partial y}\Big)^2\Big\rangle_{Av}=\Big\langle\Big(\frac{\partial w}{\partial y}\Big)^2\Big\rangle_{Av}=\Big\langle\Big(\frac{\partial u}{\partial y}\Big)^2\Big\rangle_{Av}$$

$$\Big\langle\Big(\frac{\partial v}{\partial z}\Big)^2\Big\rangle_{Av}=2\Big\langle\Big(\frac{\partial w}{\partial z}\Big)^2\Big\rangle_{Av}=\Big\langle\Big(\frac{\partial u}{\partial z}\Big)^2\Big\rangle_{Av}. \quad (4)$$

This is a questionable assumption, but using it on the measurements made in the pipe at $R=25\,000$, the dissipation summed over the cross section agreed with the total production in the section to within 10 percent. At $R=250\,000$ the dissipation was still low compared

to the production; but the difficulties near the wall were aggravated by the higher Reynolds number, and the higher speed pushed some of the dissipation into frequencies beyond the range of the differentiating circuit. In the boundary layer the dissipation summed over the thickness was too low by 17 percent. Nevertheless, the low Reynolds-number pipe results would appear to justify assumption (4) on the average.

There is more uncertainty in the point to point values of the dissipation than in the total, especially below $y^*=10$. Here the first term in expression (2) becomes significant. Because of the necessity of finding the second derivative, the uncertainty in this term is great. Some idea of its magnitude is given by the dotted curve in Fig. 8. For want of a better name it is called "dissipative transfer."

The large values of mean-square derivatives in the y and z directions in the region $y^*<100$ accounts for the improvement afforded by the more complete expression (2). For example, at $y^*\approx 10$, $\langle(\partial u/\partial y)^2\rangle_{Av}$ was about 20 times $\langle(\partial u/\partial x)^2\rangle_{Av}$ instead of twice as given by isotropy. This indicates small spatial scales of motions in the y and z directions and conveys the picture of u currents running fiber-like in the x direction. For $y^*>100$ the various mean-square derivatives were of the same order, and the assumption of isotropy improved with increasing distance. It is believed that the point to point values of the dissipation as measured in the pipe at $R=25\,000$ are fairly reliable. Those measured in the boundary layer are equally reliable for $y^*>30$. Since phenomena should be universal for the small values of y^*, say, in the neighborhood of the sublayer, the dissipation curves for the two cases are not greatly in doubt.

The question of energy balance and energy movements from place to place rests on the point-to-point values of the terms in Eq. (1). The results for the pipe at $R=25\,000$ where the information on dissipation is most reliable are given in Fig. 9. The convection term

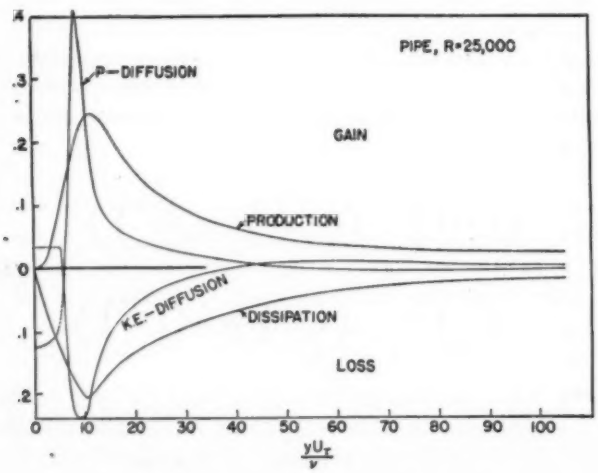


FIG. 9. Turbulent energy balance for pipe flow, wall region only.

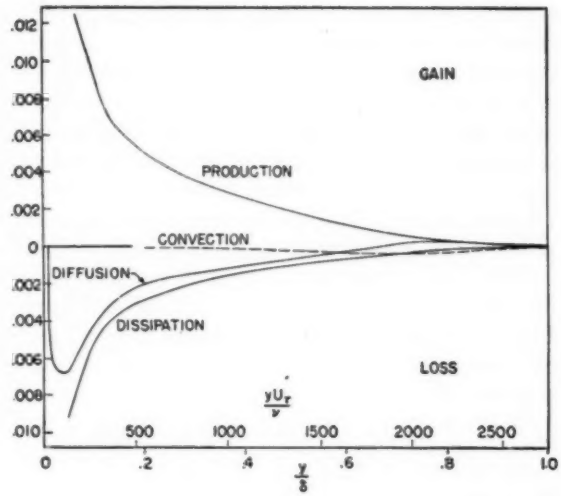


FIG. 10. Turbulent energy balance for boundary-layer flow. Diffusion is combination of kinetic and pressure energy.

is, of course, absent. The production and dissipation are closely in balance throughout the region covered by the figure and on to 0.3δ. Nearer the center the dissipation exceeds the production and at the center is about 1 percent of the maximum value. The regions gaining and losing kinetic energy by diffusion show that there is a movement of kinetic energy away from the region of high production toward the center of the pipe and possibly a small movement from the maximum toward the wall. The pressure diffusion is merely a residual of the determinations and may be greatly in error. It indicates a movement of energy toward the region containing the maxima, i.e., converging on this region from the wall and from the outside. Compared to the rates of production and dissipation, the diffusion of energy is very small.

The situation in the boundary layer appears to be different. As shown in Fig. 8, the dissipation becomes less than the production for $y^*>40$. Figure 10 shows how this deficiency persists out to $y/\delta\approx 0.65$. A difference curve representing the sum of the kinetic and pressure diffusion is shown. Since the convection term is small, the difference curve is mainly the difference between production and dissipation. Since the kinetic energy diffusion was not measured in this experiment, the two types of diffusion could not be separated. The combination represents a loss by transfer throughout the region from $0.01<y/\delta<0.65$ and a slight gain from 0.65 to 1. The gain does not balance the loss and therefore fails to yield an energy balance. Townsend's measurements[6] of the kinetic energy diffusion in a boundary layer from $0.033<y/\delta<1$ show a gain throughout. Adopting his values we must conclude the existence of loss by pressure diffusion from all points except those very close to the wall. It is not clear where the lost energy has gone. Townsend found a similar state of affairs and concluded that the energy moved toward the wall and was dissipated in the laminar sub-

layer. However, the results obtained in the pipe in the vicinity of the sublayer, which should be equally valid in the boundary layer, deny such high rates of dissipation as would be required. If we conclude that the energy moves outward, we are again unable to find where it is dissipated. In this connection it would be helpful to know more about phenomena in the intermittent region.

In spite of some uncertainty about the movement of turbulent energy, it is clear that the high production near the wall does not play the role of supplier of turbulence to the rest of the shear layer. Our best information to date shows also that the region near the wall does not play the role of energy sink for the turbulence in the rest of the layer.

5. THE SPECTRUM

If a hot-wire signal is passed through an analyzer and the mean-square output is measured, the contribution in the frequency interval n and $n+dn$ is obtained. This amounts to analyzing a time varying signal into its Fourier components. To transform from time to space, it is assumed that variations with time are caused by a nonchanging configuration being swept past a hot wire with the mean local velocity. Wavelengths in this direction are then given by U/n, and a wave number is defined by

$$k_1=2\pi n/U.$$

The spectral function may be expressed as a function of k_1. For example, $F_u(k_1)dk$, is the contribution to $\langle u^2\rangle_{Av}$ from wave numbers k_1, k_1+dk_1, and

$$\langle u^2\rangle_{Av}=\int_0^\infty F_u(k_1)dk_1.$$

This is known as the one-dimensional spectrum, so named because it represents a cut in one direction through a three-dimensional pattern. In isotropic turbulence, where the pattern is symmetrical in all directions, the function F_u bears the following relation to the spectral function of the transverse components F_v and F_w:

$$F_v=F_w=\frac{1}{2}\left(F_u-k_1\frac{dF_u}{dk_1}\right). \tag{5}$$

In an isotropic turbulence the three-dimensional function specifies the energy in scales of motion represented reciprocally by the wave number. On the basis of certain plausible assumptions,[8,9] the transfer of energy from larger scales to smaller scales may be expressed. From these come the well-known $k^{-5/3}$ law, applicable to an equilibrium range in which k is not so large that viscous dissipation enters, and the less known k^{-1} law proposed by Tchen[9] for transfer in a range influenced by mean velocity gradient. These laws are based upon the three-dimensional spectrum. Since hot wires measure only the one-dimensional spectral function, comparison between theories and available experiments require the transformation between the three-dimensional and one-dimensional functions. When the turbulence is isotropic, this transformation is known, and comparison is possible.

In shear flow, difficulties of interpretation arise from the lack of isotropy. The relation between the three-dimensional and one-dimensional spectra becomes obscured when properties depend on direction. Nevertheless, spectral measurements in a shear flow serve a useful purpose. The measurements reported in references 1 and 2 yield spectra that show equilibrium ranges where the process of energy transfer from smaller to larger wave numbers is taking place. This is generally true for $y^*>100$. Nearer the wall there is evidence of marked superposition of production and dissipation.

The spectra show one very distinctive feature that is apparently connected with lack of isotropy. This is illustrated in Fig. 11, plotted from observations in the pipe at $y/\delta=0.69$. Here the lateral spectral functions F_v and F_w are compared to the longitudinal function F_u. The solid curve is F_u and the broken curve represents F_v and F_w derived from F_u by formula (5). The two curves illustrate the difference between lateral and longitudinal functions that would exist if the turbulence were isotropic. Attention is called to the fact that the observed F_v and F_w represented by the points are

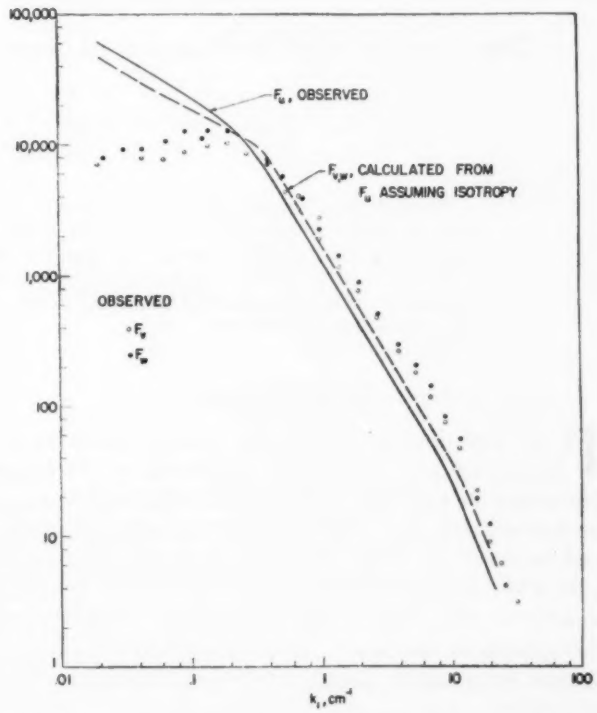


FIG. 11. One-dimensional energy spectra observed in pipe at $y/\delta=0.69$. Speed at center 100 ft per sec, $R=250\,000$.

[8] W. Heisenberg, Z. Physik **124**, 628–657 (1948).
[9] C. M. Tchen, Natl. Bur. Standards J. Research, **50**, No. 1, Report 2388 (1953).

strikingly lower than F_u at the lower wave numbers. For example, at $k_1=0.02$ the ratio is approximately $\frac{1}{8}$. Similar phenomena were observed at all positions, differing only in degree and relative value of F_v and F_w. At the center of the pipe the ratio of F_v and F_w to F_u was $\frac{1}{3}$ at $k_1=0.02$. Near the surface ($y/\delta=0.074$, $y^*=640$) F_v was lower than F_w by a factor of $\frac{1}{4}$ at $k_1=0.02$ and lower than F_u by a factor of 1/18. In the boundary layer, where only F_u and F_v were measured, a similar state of affairs existed but without the observed F_v anywhere exceeding the isotropically derived F_v. The comparative standing of $\langle u^2\rangle_{k_1}$, $\langle v^2\rangle_{k_1}$, and $\langle w^2\rangle_{k_1}$, shown in Figs. 4 and 7, is largely a reflection of the relative amounts of energy found at the lower wave numbers.

Such unusual features as this should be carefully considered for what they may possibly reveal about the character of the motions. This one may arise in part from the circumstance that longitudinal motions are greater than lateral motions. We recall that our one-dimensional spectrum is a cut through the space pattern along the direction of the component u. The pattern can be represented by waves, if we stipulate that the waves must be transverse, i.e., particle velocity parallel to the wave front. The fluctuation velocity vector then determines the direction of the wave front. When u is larger than v and w the fronts are more slanted along the u direction than when all three components are statistically equal. The projection of any arbitrary wavelength (which is always normal to the front) on the u direction is therefore greater, and the wave number is smaller. Hence, the excess of u energy over that of the lateral components automatically places itself toward the lower end of the one-dimensional spectrum. Near the wall where $\langle u^2\rangle_{k_1}$ is very much larger than either $\langle v^2\rangle_{k_1}$ or $\langle w^2\rangle_{k_1}$, the excess of energy in u motions may extend into the region of very small scales and still appear among the small wave numbers due to the fact that the wave fronts tend toward parallelism with the mean flow.

There is no obvious explanation for the apparent excess of F_v and F_w at the higher wave numbers shown in Fig. 11. The situation is somewhat confused here because boundary layer results did not show this effect, but rather a very definite approach toward isotropy in this range. Our assumed cause does not explain why the effect shown in Fig. 11 was still in evidence at the center of the pipe. It may be remarked that similar effects, to a smaller degree, have been found in turbulence behind a grid. There may then be a more subtle meaning to Fig. 11. The possibility exists that some effect associated with the transfer of energy from the u component to the lateral components is here in evidence. This rather speculative discussion illustrates some of the difficulties of interpretation that surround spectra in shear flow.

JOURNAL OF APPLIED PHYSICS VOLUME 25, NUMBER 2 FEBRUARY, 1954

The Analysis and Purification of Rare Gases by Means of Electric Discharges*

R. Riesz and G. H. Dieke
The Johns Hopkins University, Baltimore, Maryland
(Received July 13, 1953)

Through the phenomenon of cataphoresis in gas discharges, impurities are carried toward the cathode and concentrated there. This can be used for the purification of rare gases, in particular, for the removal of traces of one rare gas from another rare gas. The technique is simple and completely automatic. The same technique can be used for the analysis of rare gases. The high concentration of the impurities near the cathode increases the sensitivity by a large factor. One part of neon in 200-million parts of helium has been detected without difficulties.

1. INTRODUCTION

IT is well known that even minute amounts of impurities can change the properties of electrical discharges in rare gases. When such discharges are to be investigated, it is essential to have very pure gases and to have the assurance that the gas will remain pure while the discharge is being maintained. In order to achieve this, very careful preliminary outgassing of the system is necessary. Even then impurities often cannot be avoided because the purest commercially available gas samples may have traces of impurities present in sufficient concentration to influence the discharge characteristics, and because under the ion bombardment in a discharge, gases may be given off from the most carefully outgassed surfaces.

In order to cope with this situation, it is essential to have at one's disposal reliable methods for removing traces of impurites and, in order to know the composition of the gas, to have available methods of analysis that can detect reliably even small traces of impurities.

It is well known that for some gases the solution of the difficulties is by no means simple. It is fortunate, however, that when dealing with rare gases it is possible

* This work was carried out with support of the U. S. Office of Ordnance Research.

to solve the problem of analysis and purification in a very simple and satisfactory way. Enough experience has been accumulated in this laboratory on this method that we are confident that it can have a wide applicability.

We have reported previously[1] on the use of activated uranium as a very efficient and convenient getter for removing any impurity from a rare-gas discharge except another rare gas. As this method is very effective, we need to consider in this paper only the removal of traces of other rare gases. While this is the chief purpose of the method described below, this method will work also for the removal of other impurities although our experience is more limited along those lines.

The procedure is based on a phenomenon occurring in glow discharges known as cataphoresis, discovered as early as 1893[2] and studied experimentally by Skaupy[3] and theoretically by Druyvesteyn.[4] The treatment of Druyvesteyn was chiefly phenomenological and the experimental results of Skaupy and others fragmentary, so that the phenomenon of cataphoresis in gas discharge seems to be little known as to facts and interpretation.

While attempting to work in this laboratory with a dc glow discharge in very pure helium it was observed that the impurities were concentrated chiefly in the vicinity of the cathode, and it was recognized that this was due to the phenomenon of cataphoresis mentioned above. An extensive investigation, still in progress, has been undertaken with the aim of clearing up the mechanism of cataphoresis. This has resulted in a theory which accounts for the main features. It is, however, not the purpose of the present paper to deal with the theoretical aspects of the problem but to show how with the help of cataphoresis the practical problem of the purification of rare gases and the analysis for traces can be solved.

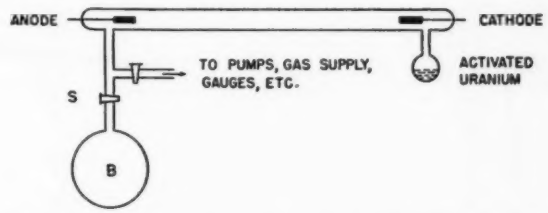


FIG. 1. Experimental arrangement for the purification of rare gases.

2. CATAPHORESIS IN A HELIUM DISCHARGE

Let us assume that we have a binary mixture of rare gases with one constituent occurring in very small concentrations. We shall take as an example helium with a trace of neon and turn later to other mixtures on which experiments have been made. Commercial "spectroscopically pure" helium is such a mixture. A

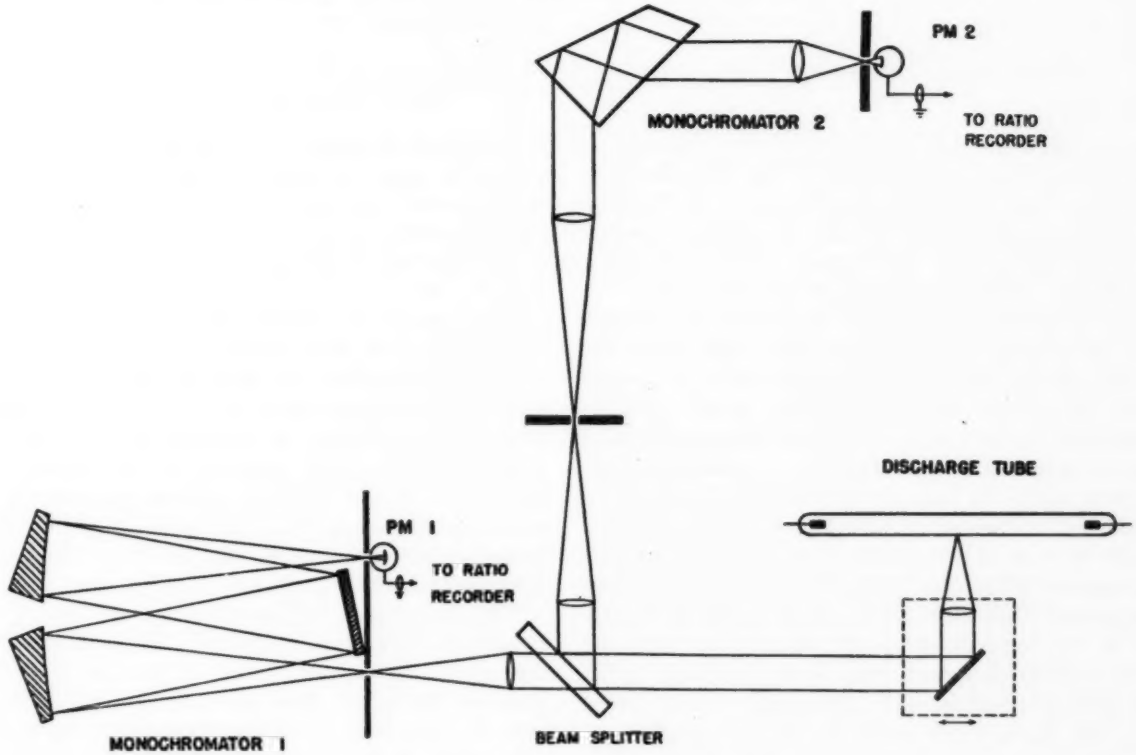


FIG. 2. Optical arrangement for ratio recording.

[1] G. H. Dieke and H. M. Crosswhite, J. Opt. Soc. Am. **42**, 433 (1952).
[2] E. C. Baly, Phil. Mag. **35**, 200 (1893).
[3] F. Skaupy, Z. tech. Phys. **6**, 284 (1925).
[4] M. J. Druyvesteyn, Physica **2**, 255 (1935).

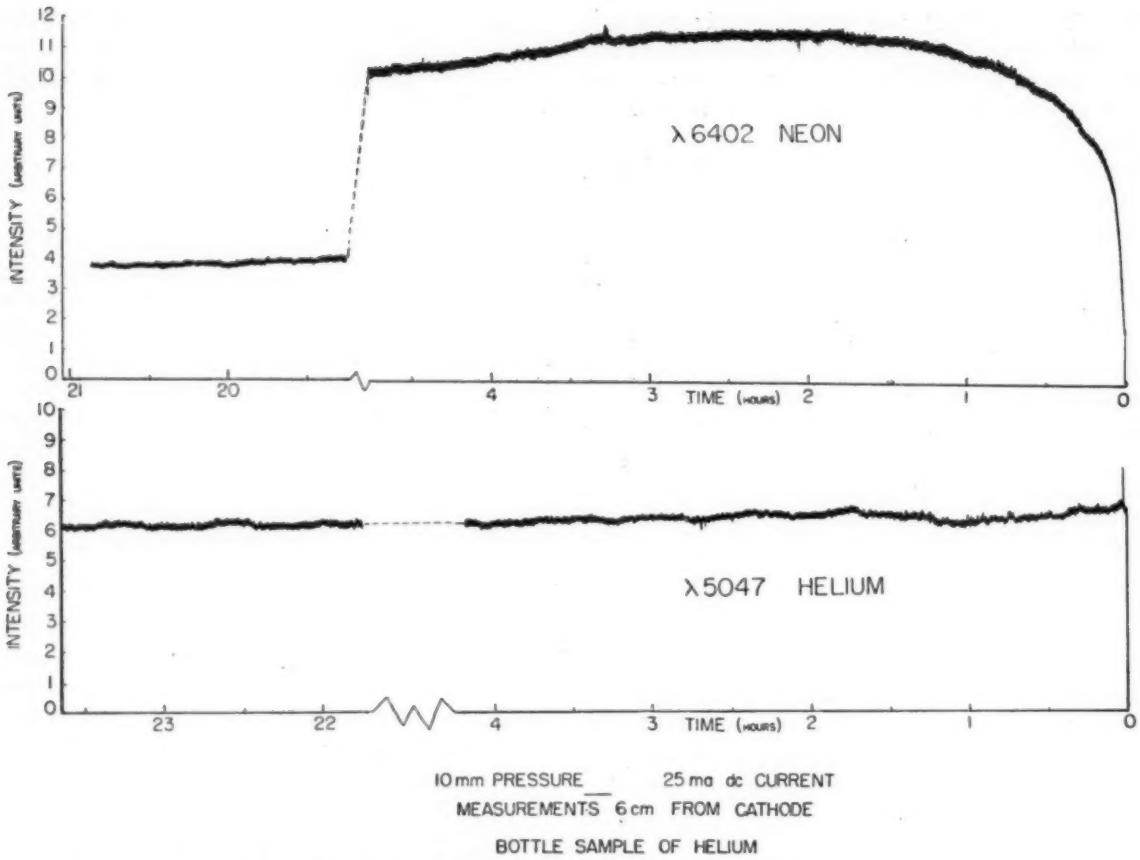


FIG. 3. Spectrum line intensity *vs* time. Commercial helium sample, 10 mm of Hg pressure, 25-ma dc current. Measurements in positive column 6 cm from cathode.

long cylindrical discharge tube is filled with this mixture to a pressure of at least 10 mm of Hg. The discharge tube (Fig. 1) has attached to it a storage bulb *B* and an activated uranium tube, as well as pressure gauges, gas supply bottles, etc. The purpose of the uranium is to eliminate all impurities that are not rare gases.

A dc discharge is maintained in the gas at currents from 10 to 30 ma. The light at any point along the axis can be fed into a spectrograph with an optical system described previously.[5] This device permits scanning along the axis of the tube. The spectrograph or monochromator is provided with a photomultiplier tube with which the intensity of any spectrum line may be recorded. Occasionally it is advantageous to record directly the ratio of two spectrum lines belonging either to the same or different elements. This is achieved with an arrangement illustrated in Fig. 2. For work with rare gases no very high demands are made on the resolving power of the spectrograph. Most work was done with a plane grating recording monochromator of 50-cm focal length designed by Crosswhite and built in this laboratory. The second monochromator was improvised by replacing the eye piece of a Hilger wavelength spectrometer with a photomultiplier tube assembly.

[5] T. Donahue and G. H. Dieke, Phys. Rev. **81**, 248 (1951).

When the discharge is first turned on with "spectroscopically pure" helium, the neon lines show up very weakly, and are uniformly distributed in intensity throughout the tube. Gradually, however, the intensity of the neon lines builds up in the vicinity of the cathode. This is illustrated in Fig. 3 which shows that the process comes to equilibrium in about two hours. During this time the intensity of the helium lines stays essentially constant. The gradual decline in the neon intensities for longer times is very probably a result of clean up of neon into the cathode. The time at which equilibrium is reached depends on the volume of the ballast bulb *B*. The larger its volume, the longer it takes before equilibrium is reached. If it is omitted equilibrium is reached within a minute, but the neon concentration near the cathode is much less.

What happens is as follows: Before the discharge was turned on there was a uniform distribution of neon throughout the whole system. When the discharge is initiated the minor constituent in the discharge tube, neon in this case, is driven toward the cathode and accumulates in the vicinity of this electrode. Neon from the storage bulb *B* then diffuses into the anode region of the discharge and is in turn pumped by cataphoresis toward the cathode. This action continues until some equilibrium distribution of neon is reached.

Figure 4 shows the distribution of neon in the tube for various pressures. At 20-mm total pressure the concentration gradient is such that the neon concentration decreases by a factor 100 along 20 cm of the tube. The decrease in concentration appears to be exponential unless the initial concentration is too high, so that for a tube length of 40 cm the concentration near the anode will be only about 10^{-4} times that at the cathode.

Although it is necessary to detect small changes in neon concentrations with quantitative measurements of line intensities, there is no difficulty in observing the qualitative effect without any instruments whatsoever. Even with "spectroscopically pure" helium when equilibrium is reached, the cathode side of the discharge is definitely red, whereas the rest of the discharge has the typically yellowish color of a helium glow discharge. The color contrast becomes quite startling if the discharge is viewed through dydymium (glassblower's) glasses which absorb the yellow helium lines. The helium discharge is then deep blue which contrasts sharply with the red neon discharge.

When the discharge is turned off, diffusion redistributes the neon uniformly throughout the entire gas volume. When the discharge is turned on again, it will appear homogeneous, and the concentration of neon at the cathode begins exactly as in a fresh gas sample. Intermediate stages may be observed when the time between turning off the discharge and turning it on again is less than necessary for complete redistribution of the neon.

3. PURIFICATION OF RARE GASES

These results suggest a very simple method for the purification of helium or any other rare gas.

The helium sample to be purified is placed in the storage bulb (B in Fig. 1) to which a discharge tube is connected as shown in Fig. 1. After the discharge is turned on it is only necessary to wait an appropriate time until the neon and any other impurity gases that might be present are pumped to their equilibrium distribution in the discharge tube. The time necessary for this depends on the amount of impurity present originally and the total volume of the gas. With a volume of 2 liters at 20 mm and such impurities as are present in commercial helium, the time is about 2 hours for a 20-ma discharge.

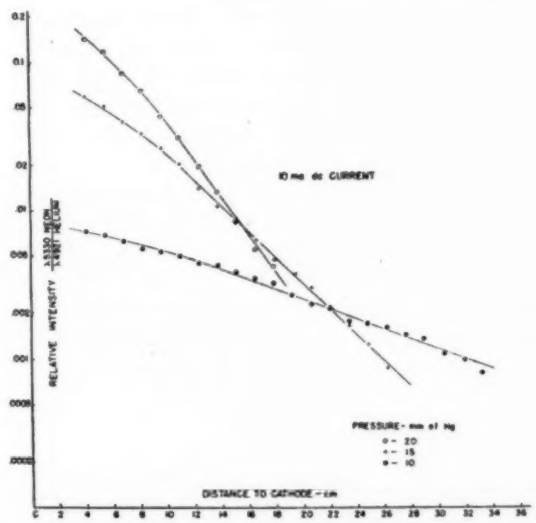


FIG. 4. Commercial helium sample. Cataphoresis as a function of gas pressure at constant current.

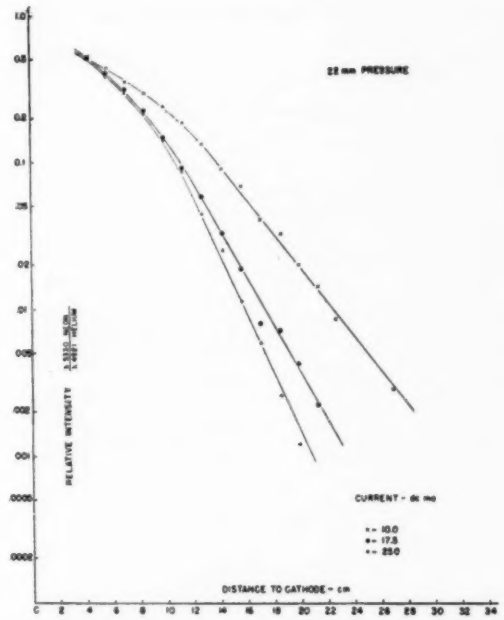


FIG. 5. Commercial helium sample. Cataphoresis as a function of current at constant pressure.

The discharge runs perfectly steadily without any supervision for months if necessary. After equilibrium is reached, the stopcock S is closed, and a highly purified sample remains in the bulb B. Should there have been too much impurity in the gas originally, one run might not be sufficient for obtaining very pure gas. The discharge tube with the impurities may then be pumped out and the process repeated. In our experience this is hardly ever necessary. The efficiency of the process is also increased by providing a larger volume around the cathode.

The values of such discharge parameters as current, pressure, and gas mixture control the magnitude of the cataphoretic action. Figures 4 and 5 illustrate how pressure and current affect the cataphoresis in the case of helium contaminated with neon. In general the slope of the neon distribution curve increases rapidly as the total gas pressure or the current is raised, indicating that the processes of gas purification and detection are greatly augmented by increases in these parameters. Experimental difficulties have limited us to pressures less than 30 mm of Hg and to currents less than 30 ma, but we believe that increases in pressure and current above these values will further enhance the effect. The previously mentioned theory accounts satisfactorily

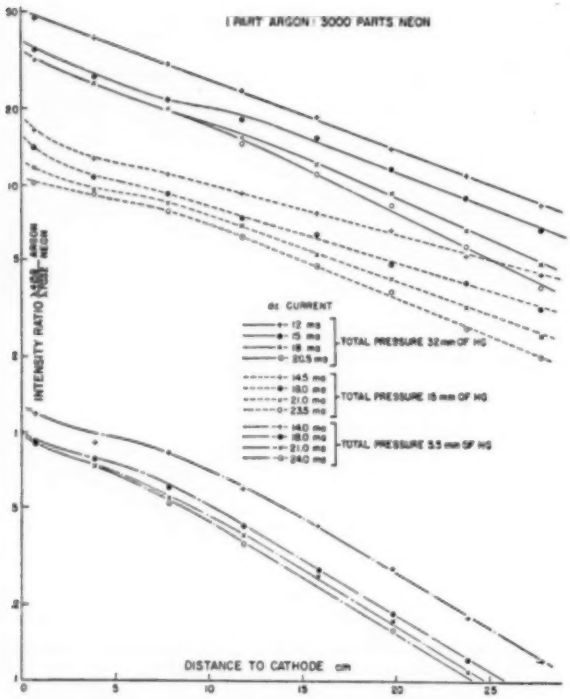


FIG. 6. Neon-argon mixture. Cataphoresis as a function of current and pressure.

for these effects on the basis of voltage gradients in the discharge and on the state of ionization of the minority gas.

We have determined that the nature of cataphoresis is such that, in the cases investigated so far, the minor constituent is removed whatever its nature. Therefore, in a mixture of krypton with traces of argon, the argon will be pumped toward the cathode whereas in a mixture of argon with traces of krypton the krypton will be removed. Moreover, the pumping action of the discharge is not restricted to other rare gases but molecular impurities such as H_2 and N_2 are removed as well. We have so far made no experiments where the *major* constituent was a molecular gas.

Binary mixtures of other rare gases have also been studied with results similar to those of He—Ne mixtures. Figures 6, 7, and 8 show the results of Ne—A, A—Kr, and Kr—Xe mixtures respectively, with the heavier gas in the minority in each case. It can be seen that the magnitude of the cataphoresis decreases as the mass of the atoms increases, indicating that the purification and detection processes become less efficient for these cases. Here several stages of purification by cataphoresis might be necessary before a satisfactory sample is obtained. We have accounted qualitatively for this lower efficiency of cataphoresis on the basis of mass ratios and ionization potentials.

We believe the simplicity of the purification of rare gases by cataphoresis and its effectiveness to be such that it can be used in preference to all other methods when gases of very high purity are required. This applies particularly when the gases are to be used for electrical gas discharges. We have used the method extensively in our laboratory and it has given excellent results. With a rather crude vacuum technique and without extensive precautions, such as careful baking, a discharge can be obtained free from impurities and kept clean, although even with extremely careful work the discharge may liberate impurities which cannot be removed easily in any other way.

The following example is perhaps typical and shows how cataphoresis can simplify the experimental procedure. In a problem on which F. T. Byrne worked in this laboratory, it was necessary to have a discharge in very pure helium. The vacuum system was quite complicated and showed a slow leak which defied discovery. This leak, which admitted only a fraction of a millimeter of air per day, made any kind of measurement impossible. An additional discharge tube of the kind shown in Fig. 1 was attached to the system so that the original discharge tube, and the rest of the system, took the place of the storage bulb *B* in Fig. 1. The discharge in the auxiliary tube was kept going continuously, and it appeared to pump all impurities out of the system. The nitrogen and oxygen of the air as well as other minor constituents were then eliminated by the activated uranium. The argon of the air accumulated near the cathode of the auxiliary tube, where soon a bluish glow developed which gradually spread over the rest of the tube. After about two months of continuous operation, there was danger that the accumulated argon might spread into the rest of the system and

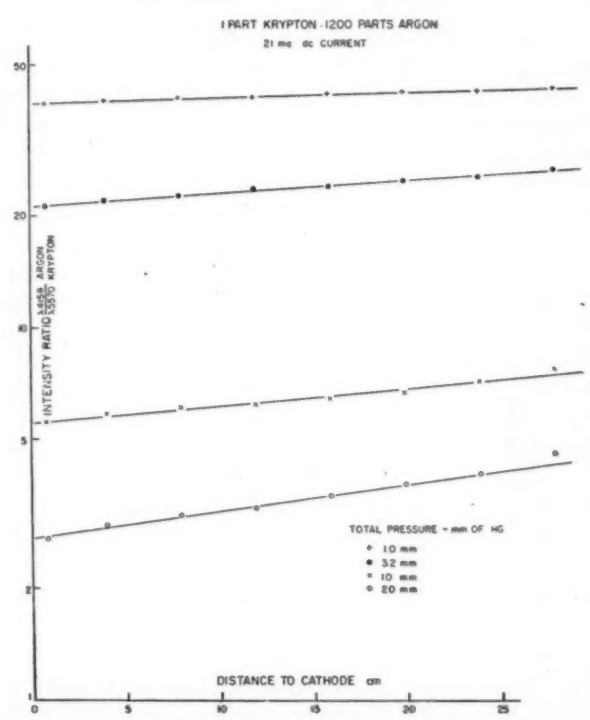

1 PART KRYPTON - 1200 PARTS ARGON
21 ma dc CURRENT
INTENSITY RATIO 4965 ARGON / 5570 KRYPTON
TOTAL PRESSURE - mm OF HG
1.0 mm
3.2 mm
10 mm
20 mm
DISTANCE TO CATHODE cm

FIG. 7. Argon-krypton mixture. Cataphoresis as a function of current and pressure.

contaminate the helium discharge. This necessitated pumping out the system and admitting fresh helium. It was thus possible to operate the helium discharge for two months without any noticeable impurities showing. Without the cataphoresis tube it would not have been possible to operate the tube satisfactorily for even a single hour.

4. SPECTROSCOPIC ANALYSIS OF RARE GASES

The same procedure that can be used for the purification of rare gases was also utilized to increase many times the sensitivity of a spectroscopic analysis of a rare gas. Let us again take the case of helium contaminated by traces of neon. The intensity of a neon line is a measure of the neon concentration. If this intensity falls below the noise level of the phototube (or below the continuous background if a photographic method is used) the limit of sensitivity is reached. We have seen that near the cathode the neon concentration can be increased many times by the cataphoretic action of the discharge, so that the neon which originally was spread over the whole volume is now concentrated in a small region. This obviously enhances greatly the sensitivity of the method.

In order to obtain quantitative results, it is necessary to calibrate the procedure. This can be done if pure gases are available which then can be mixed in known proportions. The pure gases were obtained with the method described in the previous section. It was found that within a wide range the impurity concentration was proportional to the intensity of the impurity line as compared to a line of the parent gas. A direct recording of the ratios with the arrangement of Fig. 2 greatly simplifies the procedure and eliminates errors due to fluctuations in the light intensity of the discharge. The following results on neon in helium may be of interest. Commercially available helium even when labelled "spectroscopically pure" had enough neon to make it unfit for work with pure helium discharges. On our request for information about the impurities, we received from one manufacturer the statement that mass spectrographic analysis had shown that there could be no impurity present in amounts more than 50 parts per million. Samples from other manufacturers showed no significant differences. Our analysis showed that neon was present in concentrations of 45 parts per million. With ordinary precautions we could detect about one part of neon in 200-million parts of helium. Greater sensitivity than this can undoubtedly be obtained by using a larger volume out of which to pump the impurities, by a more careful selection of the most sensitive spectral lines, and by using improved optical equipment.

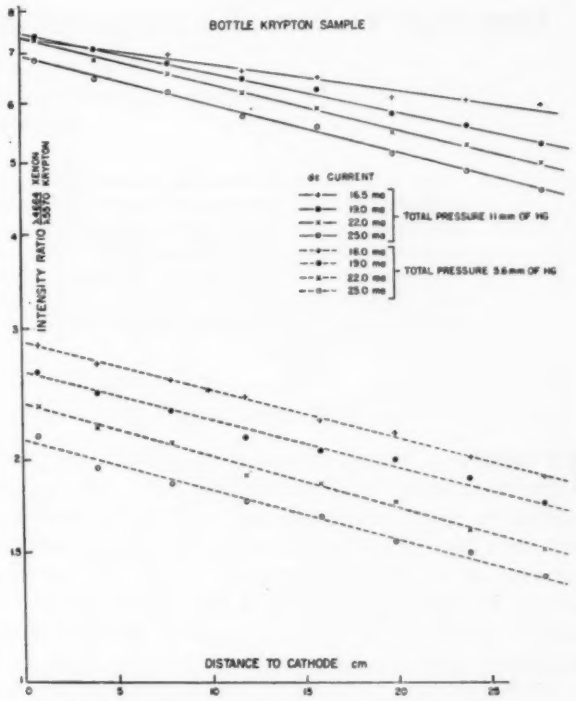


FIG. 8. Krypton-xenon mixture. Cataphoresis as a function of current and pressure.

From all that has been said it is evident that reliable results can only be obtained if the calibration is made under absolutely identical conditions as the measurements in the unknown sample. The pressure, the current, the geometry of the discharge tube, and the optical system must be kept strictly constant. As all these factors can be readily controlled, it is not very difficult to fulfill these conditions. It is furthermore important to make sure that impurities are not introduced into the discharge from the gas occluded in the electrodes through clean up during a previous use of the discharge tube. If small concentrations are involved, it is necessary to degas the electrodes carefully by induction heating or some other technique.

So far we have fewer quantitative data on the sensitivity with which other rare gas impurities can be discovered in this way. In general it is much easier to detect small amounts of a heavier constituent in a lighter gas than the reverse. This is due to the lower excitation and ionization potentials of the heavier gases. Fortunately traces of lighter rare gases rarely affect the discharge properties.

Ultra-High Vacuum. II. Limiting Factors on the Attainment of Very Low Pressures

D. ALPERT AND R. S. BURITZ
Westinghouse Research Laboratories, East Pittsburgh, Pennsylvania
(Received June 18, 1953)

Recently developed ultra-high vacuum techniques, with which it is possible to achieve working pressures of 10^{-10} mm Hg or less, have made feasible the investigation of a number of physical processes occurring at very low pressures. The Bayard-Alpert ionization gauge has been calibrated over its useful range and some of the important limitations on the production and measurement of ultra-high vacuum have been determined. It has been shown that an ultimate limitation on the achievement of very low pressures in glass systems is caused by the diffusion of atmospheric helium through the walls of the system. A simplified omegatron has been developed for the measurement of the partial pressures of residual gases in a highly evacuated system.

I. INTRODUCTION

IN recent years, numerous advances have been made in ultra-high vacuum technology.[1–4] It is now possible to produce and measure working pressures of 10^{-10} mm Hg or less in manipulative systems without the use of refrigerants, special traps, or chemical getters. In some experiments, it has been possible[2,5] to achieve conditions in which the partial pressure of the adsorbable gases has been inferred to be less than 10^{-13} or 10^{-14} mm Hg. Although the technology which has made it possible to achieve such low pressures is extremely useful in many fields of research, it has proved to be uniquely suited to the study of phenomena occurring at very low pressures.

In the course of an investigation in these laboratories of the limitation of ion gauge readings due to x-ray effects, an actual limitation in the attainment of very low pressures was observed. In a number of experiments, it was found that no matter how carefully the system had been constructed or prepared, it was not possible to achieve total pressure indications below approximately 4×10^{-11} mm Hg.[6] This was true even though flash filament observations[7] gave no indication of the presence of adsorbable gas even after weeks of operation. The present series of studies was undertaken to determine the possible limiting factors on the achievement of ultra-high vacuum.

II. RESIDUAL GAS PRESSURE IN ULTRA-HIGH VACUUM SYSTEMS

It has been shown[3] that a conventional ion gauge is limited in its ability to read very low pressures by the existence of a residual current to the ion collector. This current, which is independent of the gas density, results from x-rays formed when electrons from the filament strike the accelerating grid. The x-rays incident on the ion collector release photoelectrons which then travel to the grid, thus producing a current of the same sign as ions arriving at the collector. One can determine the portion which the x-ray effect contributes to the ion collector current by plotting a curve of this current (i_p) as a function of the voltage, V_g, applied to the grid. When the x-ray effect predominates, the log i_p *vs* log V_g curve is empirically found to be a straight line of slope approximately 1.8. At higher pressures, when the ion current is large compared to the residual current the log i_p *vs* log V_g curve is very similar to gas ionization probability curves, with a maximum between 100 v and 200 v. The Bayard-Alpert gauge significantly lowers the residual current by reducing the size of the ion collector so that the solid angle presented to x-rays is approximately 1000 times lower than for conventional gauges.

In a number of experiments designed to achieve the lowest possible pressure in a sealed-off system, the minimum ion collector current measured on the Bayard-Alpert gauges corresponded to an indicated pressure of about 1×10^{-10} mm Hg. Measurements of i_p as a function of V_g resulted in curves similar to that shown in Fig. 1. The solid curve passing through the experimental points represents the total current to the ion collector as a function of voltage, this current being the sum of the current due to x-rays and that due to gas ionization. The linear portion of the curve above 400 volts is ascribed to the dominance of the current due to x-rays, while the deviation from linearity below this voltage is considered to be due to the gas ions. On the basis of previous analyses of such curves the observed curve is considered to be the superposition of the two dashed curves below; the straight line results from the x-ray effect, while the curve below represents the gas ionization. The methods used to achieve the very low pressures are described in (I). No matter how carefully the system had been prepared or checked for leaks, the lowest pressure reading which could be achieved by these means ex-

[1] For a discussion of the work of Nottingham, Anderson, and Apker prior to 1949, see S. Dushman, *Scientific Foundations of Vacuum Technique* (John Wiley and Sons, Inc., New York, 1949) "Production of Extremely Low Pressures in Sealed-off Devices," p. 730.
[2] D. Alpert, J. Appl. Phys., **24**, 860–876 (1953), hereafter referred to as (I).
[3] R. T. Bayard and D. Alpert, Rev. Sci. Instr. **21**, 571 (1950).
[4] G. H. Metson, Brit. J. Appl. Phys. **2**, 46 (1951).
[5] J. A. Becker and W. B. Nottingham have independently reported on their ultra-high vacuum studies at the Massachusetts Institute of Technology Conference on Physical Electronics.
[6] All pressure values as reported in this paper are equivalent nitrogen pressures unless otherwise noted.
[7] L. Apker, Ind. Eng. Chem. **40**, 846 (1948).

ceeded the x-ray limit by at least 3 to 4×10^{-11} mm Hg. Since the partial pressure of the adsorbable gases was felt to be orders of magnitude less than this, the the source of this residual pressure was to be sought elsewhere.

III. LINEARITY OF THE IONIZATION GAUGE

To eliminate the possibility that the magnitude of the residual pressure reading was in doubt because of calibration errors, it was important first of all to perform an experiment to verify the linearity of the gauge over its range of usefulness. While there has been no particular question concerning the accuracy of ion gauge calibrations at very low pressures, they have never previously been checked for pressures below 10^{-6} mm Hg.

The vacuum system used in the experiment is shown in Fig. 2 and is typical of those described in (I). If at a time $t=0$, we expose one side of a constriction of conductance, S_1, to a pressure, p_A, of a given gas, the rate of rise of pressure in the volume, V_1, is given by Knudsen's laws of gas flow:

$$dp_1/dt=S_1/V_1\,p_A \quad \text{if} \quad p_1\ll p_A \tag{1}$$

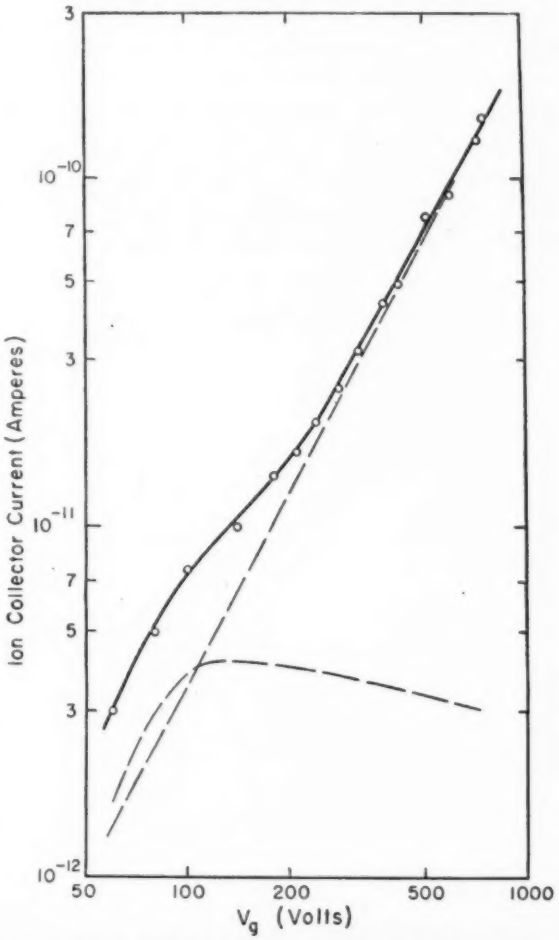


FIG. 1. Ion collector current *versus* grid voltage. Solid curve passes through experimental points. Dashed line represents x-ray effect. Dashed curve represents gas ionization current.

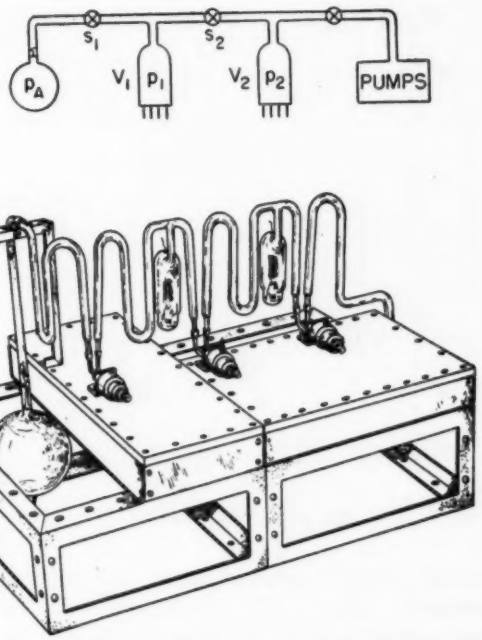


FIG. 2. Line drawing and block diagram of ultra-high vacuum system for ionization gauge linearity experiment.

so that

$$p_1=at+p_{10}, \tag{2}$$

where $a=(S_1/V_1)p_A$ and p_{10}=the residual pressure in V_1. A linear increase of the ion gauge reading with time constitutes a check on the linearity of the gauge. This method has, in fact, been used by Dushman and Found[8] to indicate the linearity of a gauge above 10^{-6} mm Hg. If one were to try to extend this method down to 10^{-10} mm Hg, it would be necessary to adjust the conductance, S_1, so that the rate of rise in V_1 would be sufficiently small ($\simeq5\times10^{-10}$ mm Hg per minute) to give enough time to take meaningful readings between 10^{-10} and 10^{-9} mm Hg. With such a small rate of rise, however, it would take about three years to cover the entire range of the gauge. Hence, a second valve[9] with conductance, S_2, and an ionization gauge were added as shown. The rate of rise of p_2, the pressure in volume V_2, is now given by

$$\frac{dp_2}{dt}=\frac{S_2}{V_2}\,p_1=\frac{S_2}{V_2}(at+p_{10}). \tag{3}$$

For $t\geq1$ second $p_{10}\ll at$. The solution of (3) is then of the form

$$p_2=p_{20}+bt^2 \tag{4}$$

where p_{20} is the residual pressure in volume V_2 and b is a constant of proportionality. Thus the pressure in V_2 is expected to rise quadratically in time, and it is possible to take readings at very low pressures with a

[8] S. Dushman and C. G. Found, Phys. Rev. **15**, 7 (1921).
[9] The constrictions used were standard all-metal vacuum valves. D. Alpert, Rev. Sci. Instr. **22**, 536 (1951).

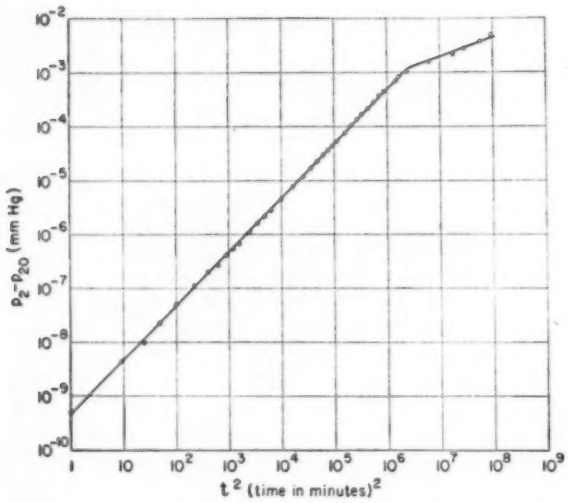


FIG. 3. p_2-p_{20} *versus* t^2 for argon.

low rate of rise and still cover the entire range in a reasonable time. A plot of p_2-p_{20} *vs* t^2 should thus give a straight line if the gauge is linear.

The experimental data are plotted on log-log scale in Fig. 3. The results are given for a calibrating run with argon, but similar curves have been obtained for helium. The electron current from filament to grid was 10 ma. The break in the curve above 10^{-3} mm Hg was found to coincide with the initiation of a visible glow discharge between the elements of the tube. Readings were taken by intermittently applying the operating voltages to the gauge (a reading was taken in two to three seconds) and were automatically recorded on a Leeds and Northrup Speedomax instrument. The data obey the law of Eq. (4) for almost seven decades of pressure, indicating that the gauge is linear over this range.

IV. VAPOR PRESSURE OF TUNGSTEN

Having ascertained the linearity of the gauge, we proceeded to investigate various possible sources of the residual pressure. Among the questions asked were: What is the ion gauge reading due to the finite vapor pressure of tungsten at the filament operating temperature? On the basis of available data,[10] a computation shows that at this temperature (2300°K) approximately 5×10^{10} atoms per second leave the filament in all directions. From this rate of evaporation[11] one can calculate the approximate number of tungsten atoms within the grid volume at a given instant. If information were then available on ionization probabilities for tungsten (this to our knowledge has never been measured) it would be possible to estimate the equivalent nitrogen pressure due to the tungsten atoms leaving the filament and flying through the grid volume on their way to the glass envelope. A calculation was made on the assumption that the ionization probability for tungsten is comparable to that of mercury. The resulting estimated value for the residual ion gauge reading due to the finite vapor pressure of the tungsten filament was 10^{-12} mm Hg.

The Bayard-Alpert gauge made possible a simple experiment to check this estimate. The "spare" filament in the gauge was set at a positive potential to prevent the emission of electrons, while its temperature was varied over a wide range. The regular filament was used in normal fashion to read the equivalent nitrogen pressure of tungsten atoms in the gauge as a function of this temperature. The entire ion gauge was immersed in liquid nitrogen during the run to eliminate effects due to heating of the glass envelope.

The experimental points were taken for filament temperatures between 2600°K and 3300°K. Figure 4 shows the data plotted in the manner usual for reaction rates, the logarithm of the pressure *versus* the reciprocal of the absolute temperature. As shown, the data fall on a straight line over three decades of pressure, so that it is quite reasonable to extrapolate down to the normal operating temperature of the filament. The extrapolated value at 2300°K is very close to the estimated value of 10^{-12} mm Hg. To verify that the effect is due to the vaporization of tungsten rather than to extraneous heating effects, we note that the slope of the curve should be characteristic of the metal and is approximately equal to the heat of vaporization of tungsten. The value thus obtained for the slope is 41 600° which agrees within 0.7 percent with the corresponding value obtained by Riemann.[10]

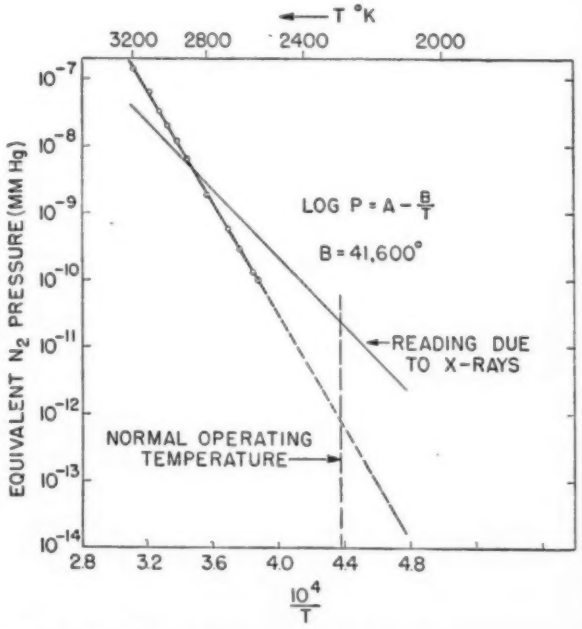


FIG. 4. Equivalent pressure due to tungsten vapor *versus* reciprocal temperature (curve through experimental points). Also plotted is the current due to x-rays which is proportional to the total electron current from the filament.

[10] A. L. Reimann, Phil. Mag. **27**, 834 (1938).

[11] This rate is so small that it takes several days to form a monolayer of tungsten on the surface of the glass envelope.

From the above discussion, an eventual low pressure limit of 10^{-12} mm Hg is set by the operation of the filament at 2300°K. With the present gauge design this is small compared to the residual reading due to x-rays, as shown in Fig. 4. Since it is possible to reduce the x-ray effect by changes in gauge geometry, it may be desirable in future gauge designs to operate the filament at a lower temperature to reduce the reading due to tungsten ions. In this case the ultimate limitation on the low pressure reading may be determined by the value of the lowest ion collector current which it is possible to measure.

Thus the ultra-high vacuum techniques have provided a novel experimental approach to studies of vaporization or sublimation of metals of low vapor pressures. However, the discussion indicates that the finite vapor pressure of the ion gauge filament is not the source of the residual pressure reading in our vacuum systems.

V. DIFFUSION OF HELIUM THROUGH GLASS

Information concerning the possible source of the residual gas was first obtained from measurements of the rates of pressure rise in sealed-off vacuum systems. It was observed that there remains a residual rate of rise of pressure in a vacuum system isolated from all pumping action, no matter how carefully it has been prepared or checked for leaks. For a typical pressure-time curve as shown in Fig. 5, the value of dp/dt is 2.88×10^{-11} mm Hg. The slope varies among different systems depending on their volume-to-surface ratio. The rate of rise has been measured and shown to be constant for over 75 days. Even when an appreciable pressure of gas had built up, no flash filament effect

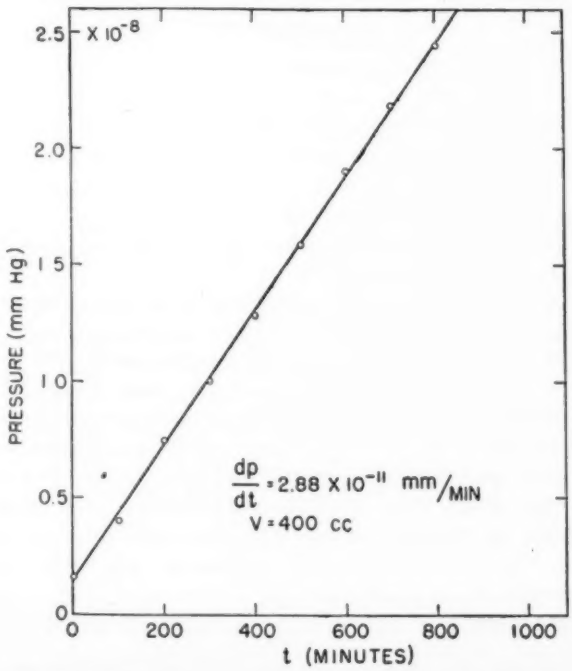


FIG. 5. Rate of rise of pressure in a typical vacuum system.

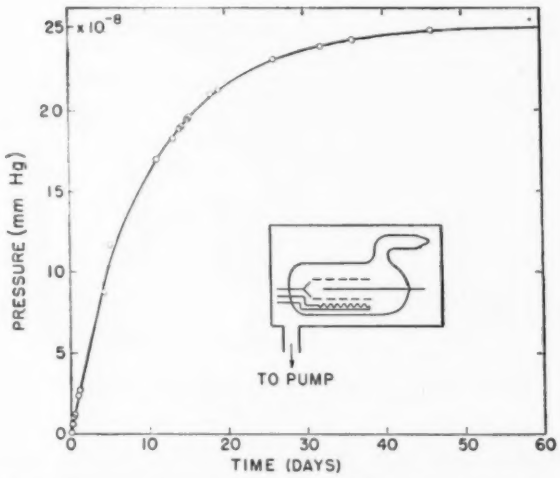


FIG. 6. Pressure *versus* time. Sealed-off ionization gauge in surrounding vacuum. External system evacuated at time $t=0$.

was observed. From these observations, it was inferred that an inert gas, probably a noble gas, comes from outside the enclosure, possibly diffusing through the glass itself. From available data on permeation rates the value of the measured rate of rise was shown to be consistent with the hypothesis that helium, present in the atmosphere to only a few parts in a million, was diffusing through the glass walls of the enclosure.

To check this hypothesis, a simple experiment was proposed to check whether the gas came from outside the glass envelope and to determine, if possible, the mechanism involved. The experiment consisted of placing a carefully sealed-off ionization gauge (vol. 230 cc, surface area 200 cm²) within a vessel which could itself be evacuated. The rate of rise in the gauge was initially measured in a normal atmosphere; then at a given time the surrounding vessel was evacuated. For an external gas source, the experiment should indicate a disappearance of the rate of rise. Figure 6 shows the pressure in the sealed-off ionization gauge plotted as a function of time after the external vessel was initially evacuated. It is seen that the rate of rise did, indeed, disappear but that it took over 60 days to reach this state of equilibrium. The external origin of the gas was thus demonstrated. Furthermore, the rate at which equilibrium was established strongly indicated a diffusion process. Detailed consideration of the experiment showed that one could get not only the equilibrium permeation rate from the initial rate of rise in Fig. 6, but also from the remaining portion of the curve, values for the diffusion coefficient and solubility of the gas in glass. Analysis of the data resulted in values for these constants which we should have liked to compare with known values for helium in glass. The literature[12] indicated, however, that the

[12] F. J. Norton, J. Am. Ceram. Soc. **36**, 90 (1953); T. F. Newkirk and F. V. Tooley, J. Am. Ceram. Soc. **32**, 272 (1949); W. D. Urrey, J. Am. Chem. Soc. **54**, 3887 (1932).

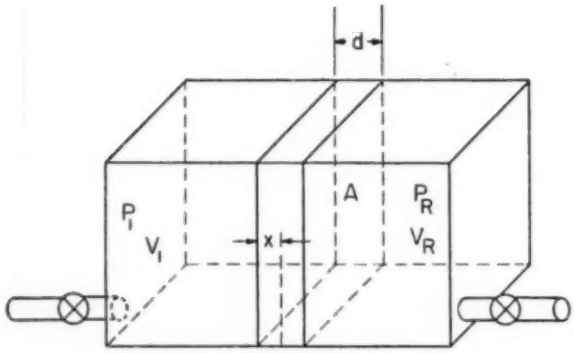


FIG. 7. Block diagram of helium diffusion experiment. Membrane area A, thickness d.

diffusion coefficient, D, had never been directly measured and that the solubility, S, had been measured[13] only for one temperature (515°C). Most previous workers[12,13] have measured only the permeation rate, which corresponds to the product of D and S.

A new experimental program was therefore proposed. It was felt that the new ultra-high vacuum techniques provided an effective means for studying the permeation rate, k, as well as the diffusion coefficient, D, and the solubility, S, for helium in glass. The ability to measure extremely small rates of pressure change made it feasible to measure gas flows at lower temperatures than were previously possible and to eliminate many of the sources of error.

The detailed results of the experiments carried out will be published shortly[14]; however, for the purpose of completeness, certain of the results will be quoted here.

The experiment was quite simple in principle. As shown in the block diagram of Fig. 7, the apparatus consisted of two independent vacuum systems separated by a Pyrex glass wall. The manifolds contained valves for isolating the systems from the pumps, as well as a helium gas source and ionization gauges for measuring pressures. The entire apparatus was initially baked at high temperatures and evacuated to very low pressures ($\sim 10^{-9}$ mm Hg). At time $t=0$, helium at pressure p_1 was introduced into the volume on one side of the membrane; thereafter the pressure p_R in the receiving volume V_R was periodically measured. A typical run at room temperature is shown in Fig. 8. As one would expect in a diffusion process, the curve starts upward from zero slope, the rate of rise of pressure gradually increasing until a steady state is achieved. It is seen that at room temperature equilibrium has not quite been achieved even after 10^4 minutes. To evaluate the diffusion constants, it is necessary to fit the experimental data to a suitable mathematical solution of the diffusion equation.

The current density $I(d,t)$ of gas atoms passing through the membrane is given for the case of a plane membrane of uniform thickness d by the diffusion law,

$$I(d,t) = -D \left. \frac{\partial n(x,t)}{\partial x} \right|_{x=d}, \tag{5}$$

where $n(x,t)$ is obtained from the time-dependent solution of the diffusion equation

$$\partial n/\partial t = D\partial^2 n/\partial x^2 \tag{6}$$

with the appropriate boundary conditions.

An exact solution of the above equations is possible involving the summation of an infinite series of exponential terms. It is given here without proof to illustrate the dependence of pressure on the various

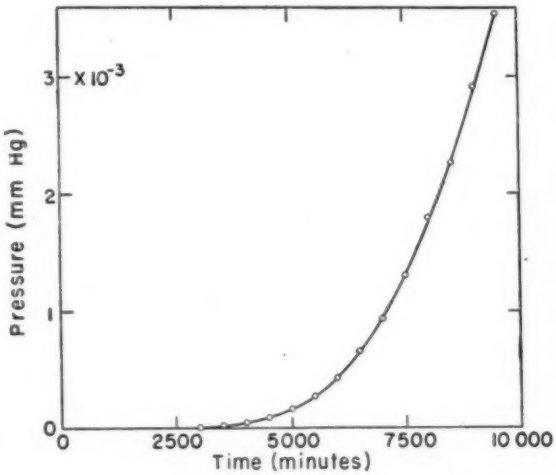


FIG. 8. Receiver pressure *versus* time. First week.

diffusion parameters. The pressure p_R in the receiving volume is given as a function of time by the expression[15]

$$p_R = \frac{A d p_1}{6 V_R} S \left\{ \frac{6Dt}{d^2} - 1 + \frac{12}{\pi^2} \sum_{m=1}^{\infty} \frac{(-1)^{m-1}}{m^2} \exp\left(-m^2 \frac{\pi^2}{x} \frac{Dt}{d^2} \right) \right\},$$

where A and d are the membrane area and thickness, respectively; S is the solubility of helium in the glass; and D is the diffusion coefficient of helium in glass.

In principle it is possible to obtain both D and S by a direct comparison of the experimentally observed curve with the above solution; in practice it is necessary to use more tractable mathematical solutions. For example, if one waits long enough (i.e., for t sufficiently large) the exponential terms become negligible and one has the remaining linear expression describing the asymptotic equilibrium rate of pressure rise. This constitutes one way of carrying out the experiment.

[13] G. A. Williams and J. B. Ferguson, J. Am. Chem. Soc. 46, 635 (1924).

[14] Rogers, Buritz, and Alpert, "Diffusion coefficient, solubility, and permeability of helium in glass," J. Appl. Phys. (to be published).

[15] T. Holstein, "Solution of a Diffusion Problem," Westinghouse Research Report R-94411-D.

However, Fig. 8 demonstrates that it would require several weeks to accumulate sufficient data for a single run at room temperature. As indicated in the detailed discussion,[14] it was necessary to utilize other forms of the solution to make possible the evaluation of the diffusion constants over a large temperature range.

From these experiments, values were obtained for the diffusion coefficient and solubility of helium in glass. The values at room temperature are listed in Table I.

For the purpose of comparison, we also list the corresponding values for the "unknown" gas in the glass walls of the ion gauge, as similarly deduced from the data of Fig. 6. The solubility is calculated using the known partial pressure of helium in the atmosphere.

The agreement in the measured values of both the diffusion coefficient and the solubility gives strong confirmation of the hypothesis that the residual rate of rise is due to the diffusion of atmospheric helium into the vacuum systems in question.

VI. MEASUREMENT OF PARTIAL RESIDUAL PRESSURES

As a result of the helium-diffusion studies, it became clear that for certain researches requiring ultra-high vacuum, a new approach to the measurement of pressures was desirable. The ionization gauge gives a measure of the total residual pressure in a system. In many experiments, it is more important to know the partial pressures of some of the residual gases rather than the total pressure due to all of them. For example, in many experiments involving clean surfaces, a small partial pressure of helium would not be expected to affect the results adversely. This is also true for some investigations in gaseous electronics. Hence a need was established for an ionization gauge, or rather, a mass spectrometer for the measurement of the partial pressures of the individual gas constituents.

The instrument we have developed to serve this purpose is a simplified version of the "omegatron."[16] It is simple in structure and operation and is capable of high sensitivity and reasonable resolution. It can withstand bake-out at high temperatures and thus permits the attainment of very low pressures.

The principle of operation of the omegatron, as shown schematically in Fig. 9, is similar to that of the cyclotron. A narrow cylindrical beam of ionizing electrons is formed parallel to the direction of a magnetic field. Ions formed along the beam are accelerated by the alternating potential between two flat rf plates which correspond to the dees of a cyclotron. When the value of the impressed r f frequency is equal to the cyclotron frequency $\omega = eH/M$, the ions of mass M and charge e are accelerated in orbits of increasing size and eventually strike an ion collector. The ion collector current is a measure of the abundance of the gas of this charge to mass ratio. For a magnetic field of 2100 gauss, as used in some of these experiments, the calculated cyclotron frequency is 3.2 megacycles for mass 1 and 32 kilocycles for mass 100.

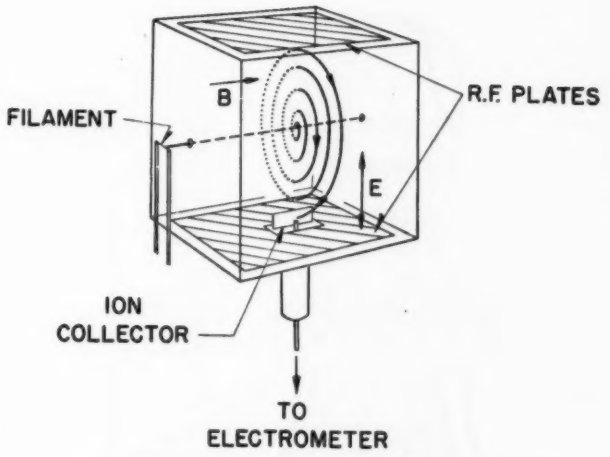


FIG. 9. Schematic diagram of simplified omegatron.

A drawing of the simplified omegatron is shown in Fig. 10. The rf plates and the surrounding shield box are fabricated in the form of a cube from 0.015-inch Advance (copper-nickel alloy) sheet. The electrons are accelerated through holes $\frac{1}{16}$-inch in diameter in opposite walls. The ion collector, also fabricated from 0.015 Advance material, projects through a slit $\frac{1}{16}$-inch wide a distance of approximately $\frac{1}{16}$-inch into the accelerating volume. The conducting lead from the ion collector is brought out through an Inconel coaxial shielding cylinder. As shown, an external shielding cylinder completes the connection to the electrometer which measures the ion current.

In practice, omegatrons have been made in two sizes, one with a shielding box in the shape of a 2-cm cube, the other a 1-cm cube. The data presented here were taken with the larger instrument. The impressed rf voltage (of the order of one volt in amplitude) was supplied by a Hewlett-Packard Model 650A test oscillator, whereas the ion current was measured with a vibrating-reed electrometer. Depending somewhat on the operating conditions, it was sometimes desirable to apply a small negative "trapping voltage"[15] of a volt or less between the rf plates and shielding box.

The operational characteristics of the omegatron were investigated by attaching one of the instruments to an ultra-high vacuum system. Since it was desired to observe a number of peaks, the system was de-

TABLE I. Measured values of D and S.

	Helium in Pyrex	"Unknown gas" in walls of sealed-off ion gauge
T°C	27°	27°
D (cm²/sec)	7.74×10^{-9}	1×10^{-8}
S (cc at STP/cc×atm)	1×10^{-2}	2×10^{-2}

[16] Sommer, Thomas, and Hipple, Phys. Rev. 82, 697 (1951).

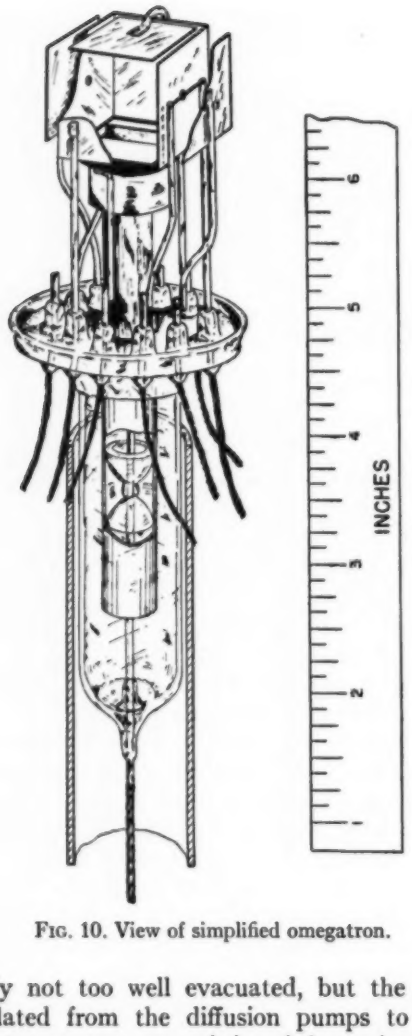


FIG. 10. View of simplified omegatron.

liberately not too well evacuated, but the manifold was isolated from the diffusion pumps to make it possible to measure rates of rise of the various gases.

A typical curve of ion collector current as a function of impressed rf frequency is shown in Fig. 11. In this experiment, the total pressure in the system as read on an ionization gauge was 10^{-7} mm Hg. The ionizing electron current through the omegatron was 4 microamps, while the accelerating voltage was 90 volts and the applied rf voltage approximately 2 volts. The magnetic field, supplied by a permanent Alnico magnet, was 2100 gauss. It is seen that in a general way the predicted frequencies for the various e/M ratios were experimentally verified. Since the ion accelerating region within the omegatron does not provide simple field configurations, it is not surprising that there is some discrepancy between the observed and the calculated frequency for a given mass. There is a systematic deviation from a linear relation between mass number and frequency which cannot be attributed to the lack of accuracy in the knowledge of the magnetic field. On the other hand, the instrument is quite capable of adequate resolution for mass numbers up to at least mass 40. By alternately introducing helium and nitrogen into the system, the resonant frequencies corresponding to masses 4, 14, and 28 were identified positively.

It is interesting to observe that the sensitivity of the omegatron is comparable to that of a conventional ionization gauge. In Fig. 11 the helium peak alone is approximately 2.5×10^{-12} ampere for an electron current of 4 microamperes at an ion gauge pressure of 10^{-7} mm Hg. A simple calculation gives a value for the sensitivity of approximately 10(mm Hg)$^{-1}$ as usually defined[2] for ion gauges. This is almost the same as the value for the sensitivity of the Bayard-Alpert gauge. Since the average distance traversed by an electron is comparable in the two instruments, it follows that a large fraction of the ions formed in the omegatron must arrive at the ion collector.

The omegatron provides a unique tool for investigating the source of the residual rate of rise in sealed-off vacuum systems by measuring the variation in height of the observable peaks in a completely isolated system. Figure 12 shows data for peak height as a function of time for such a system. The linear increase of the mass 4 peak with time is conclusive evidence that the residual rise in pressure is largely due to helium. The other mass peaks, principally mass 28, show no observable variation of peak height with time.

If the mass 4 peak height is measured as a function of the total pressure as read on an ionization gauge, one has a further check on the linearity of the omegatron. Figure 13 shows a plot of such measurements taken over a pressure range from 10^{-9} to 10^{-7} mm Hg. The corresponding values of the ion current from the omegatron fell between 2×10^{-13} amp and 2×10^{-11} amp for an electron current of 20 microamperes. It will be noted that the intercept of this curve on the pressure axis should give the partial pressure contributed by non-helium gases in the system. While it is somewhat difficult to establish the intercept accurately, it was demonstrated to be small compared to

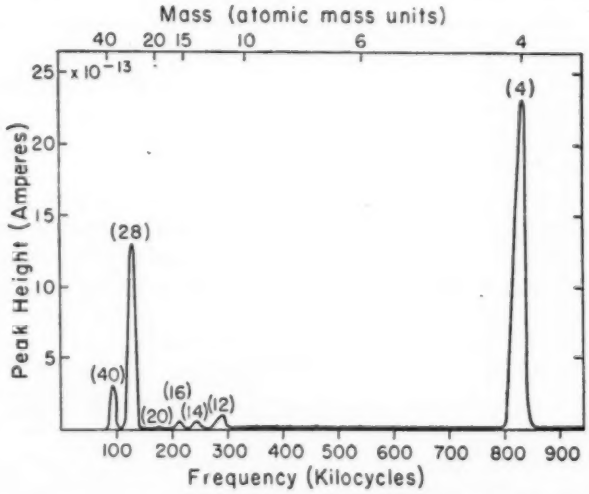


FIG. 11. Ion collector current *versus* impressed rf frequency. The mass scale above has been adjusted to coincide at the frequency of the mass 4 peak.

the total residual pressure in the system which in this case was approximately 10^{-9} mm Hg. Thus we have a final confirmation of the nature of the gas diffusing into a sealed vacuum system.

VII. QUANTITATIVE MEASUREMENT OF LOWEST ACHIEVABLE PRESSURE

It has been indicated that the lowest attainable equivalent nitrogen pressure was approximately 4×10^{-11} mm Hg. This was achieved by the methods described in (I) with the vacuum system sealed from the pumps and maintained at the above pressure by the continuous pumping action of the ionization gauge. It has also been shown that the residual gas is most probably helium, which diffuses into the system at a known rate. The question which remains is: Is the quantitative value of the minimum pressure consistent with the values for the rate of influx of gas together with the pumping speed of the gauge?

In (I) it was shown that the pumping action of an ion gauge for noble gases resulted from ion removal. For this mechanism, the rate of change of pressure, p, in a sealed-off system of volume V is given by

$$dp/dt=-(S/V)(p-p_u), \tag{7}$$

where S is the pumping speed for this mechanism, and p_u is the helium pressure at which the gauge ceases pumping. If we assume for the moment that this ultimate pressure is small compared to the pressures which have been achieved, we may set $p_u=0$.

If there is a constant rate of influx of gas diffusing into the system, the above expression becomes

$$dp/dt=-(S/V)p+C \tag{8}$$

where C is equal to the rate of pressure rise in the system in the absence of any pumping action. The solution of (8) is

$$p=C\tau+(p_0-C\tau)e^{-t/\tau} \tag{9}$$

where $\tau=V/S$.

Equation (9) describes the pressure as a function of time under the combined action of ion pumping and a constant influx of gas. The equilibrium value for large t is given by $p_E=C\tau$.

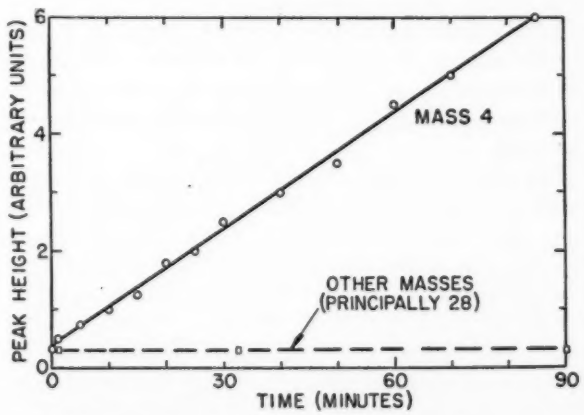


FIG. 12. Peak heights *versus* time, sealed-off system.

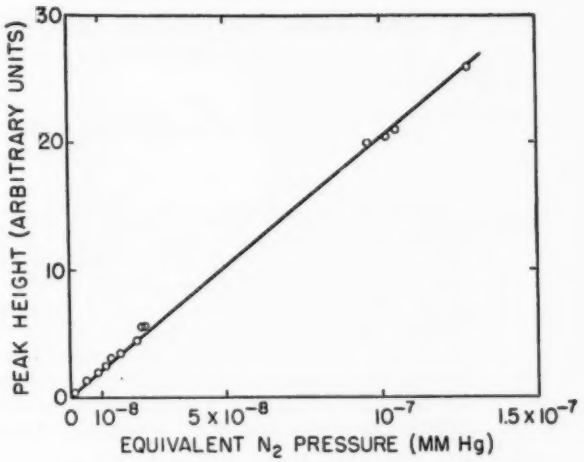


FIG. 13. Mass-4 peak height *versus* equivalent nitrogen pressure (mm Hg) as read on ionization gauge.

It was shown in (I) that the value of τ is expected to be of the order of one minute. Experimental measurements lie between 1 and 3 minutes. From Fig. 5, C is approximately equal to 3×10^{-11} mm Hg per minute, so that the product $C\tau$ is approximately 3 to 9×10^{-11} mm Hg. The measured equilibrium pressure of about 4×10^{-11} mm Hg is thus generally consistent with the values for the permeation of helium through glass and the pumping speed of an ionization gauge for helium.

VIII. CONCLUSIONS

To summarize the investigations which have been described, it has been possible to establish the linearity of an ionization gauge over most of its useful range (5×10^{-10} to 10^{-3} mm Hg). The effect of the finite vapor pressure of the tungsten filament has been measured by an experimental method which may be generally useful for the measurement of the vapor pressure and heat of vaporization of such materials. A new approach has been proposed for the experimental investigation of the diffusion of gases through solids. As a result, the permeation rate, diffusion coefficient, and solubility of helium in Pyrex glass have been measured. The predominant source of residual gas in a highly evacuated system has been shown to be atmospheric helium. A new instrument has been developed for measuring the partial pressures of the residual gases in a vacuum system.

ACKNOWLEDGMENT

We wish to thank our many colleagues for their help and advice in these investigations. In particular, we wish to acknowledge the help of C. G. Matland, W. A. Rogers, and P. R. Malmberg in the studies of diffusion through glass and D. J. Grove and J. H. Carmichael for their important contributions in the design and operation of the omegatron.

Fast Time Analysis of Intermittent Point-to-Plane Corona in Air I. The Positive Point Burst Pulse Corona*

M. R. AMIN
Department of Physics, University of California, Berkeley, California
(Received July 8, 1953)

Techniques for fast oscilloscopic analysis of intermittent point-to-plane coronas in air are described. Such analysis shows the positive point burst pulse corona to consist of an initial pulse followed by secondary pulses at varying time intervals. The duration of both pulses is less than 10^{-7} second, the initial pulses having from 1 to 2×10^{7} ions and the secondaries from 1 to 5×10^{6} ions depending on the voltage. Time intervals between pulses range from the clearing time of ions across the gap (milliseconds) to random intervals as short as 2 microseconds for secondary pulses well above threshold. Thus, in analogy to Geiger counter terminology dead times run from 2 microseconds up, and clearing times are in the order of milliseconds corresponding to local clearing of space charge near the point for secondaries, and removal of massive space charge from the low field region of the gap. A complete sequence of events from threshold to steady corona is analyzed, and indicates the predominance of self-counting through negative ion formation in air.

INTRODUCTION

THE intermittent discharges in point-to-plane geometry have been the subject of study in this laboratory for the past fifteen years under the direction of L. B. Loeb. The pioneering work was done by Trichel[1, 2] and Kip.[3, 4] Since then much more understanding of these processes has been made. In recent years, the work of English[5, 6] and the systematic studies of Bandel[7] have produced adequate data for the current-voltage characteristics and various onset values. However, the data on dynamic behavior of these discharges, due to inadequate instruments of the past, have been quite meager. In fact, except for the one micro-oscillogram of English[8] for the positive pre-onset streamer and the negative Trichel pulse, no other data suitable for a time scale analysis were available.

In this work a time study of these processes was made by using a photomultiplier and correlating the resultant signals with the electrically induced pulses, obtained from a resistor in series with the plane. Three general investigations were carried out:

1. A study of the details of the pre-onset positive point burst pulse corona.
2. A study of pre-onset positive point corona streamers.
3. A study of Trichel pulses from negative points.

Each of these will be discussed separately in Parts I, II, and III of this article.

APPARATUS AND TECHNIQUES

The power supply used in this work was built in this laboratory on the basis of a stabilized circuit proposed by Yarmotz and Pepinsky.[9] The dc output voltage of this instrument varied smoothly from zero up to 16 kv. The voltage could be measured to within ±20 volts. The maximum current output of the power supply was 10 ma. Positive and negative potentials were available from the power supply.

Two types of photomultiplier tubes were available: RCA types 1P21 and 1P28. The former was generally used because of its low dark current and noise level. The power supply for the photomultiplier was a series tube regulated type. Its voltage could be smoothly varied from −500 volts to −1500 volts.

The photomultiplier was placed in a housing unit and was mounted on a support which could be moved both vertically and horizontally. A lens system focused the region of gap under study on a rectangular slit placed in front and very close to the photomultiplier. Thus, by varying the width or the height of the slit, it was possible to study any portion of the gap desired. The slit could be made so small that a 0.5-mm×0.5-mm section of the gap was easily focused on the photocell.

The signal from the photomultiplier was sent through a terminated 200-ohm coaxial cable either directly or through a preamplifier to a synchroscope. Care was taken to match all the cables so that the study of very short pulses (of the order of 10^{-8} second) was possible.

A bell jar which could be sealed to a glass plate served as the chamber. The inside of the jar was screened with fine mesh wire to avoid charging of the glass. A protective screen was used outside the glass. Both screens were grounded.

The plane electrode was made of brass and had a radius of 7 cm. This limited the work to the relatively short gaps where the ratio of the radius to the gap distance was more than unity. Various sizes of hemispherical platinum points were used with the exception of a 2.5-mm diameter point which was brass.

A screen electrode with auxiliary field was used to screen the plane from inductive pulses, and in a manner

* This work was supported for one year on funds from the U. S. Office of Naval Research and for a second year under a fellowship from the Research Corporation of America.
[1] G. W. Trichel, Phys. Rev. **54**, 1078 (1938).
[2] G. W. Trichel, Phys. Rev. **55**, 382 (1938).
[3] A. F. Kip, Phys. Rev. **54**, 139, (1938).
[4] A. F. Kip, Phys. Rev. **55**, 549 (1939).
[5] W. N. English, Phys. Rev. **71**, 638 (1947).
[6] W. N. English, Phys. Rev. **74**, 170 (1948).
[7] H. W. Bandel, Phys. Rev. **84**, 92 (1951).
[8] W. N. English, Phys. Rev. **77**, 850 (1950).
[9] R. Pepinsky and P. Yarmotz, Rev. Sci. Instr. **19**, 247 (1948).

similar to that of Kip,[4] this was used for measuring the number of ions produced as well as the time of ion crossing. The dc voltage for the screen was provided by a bank of batteries and a potential divider which gave smooth variation.

The plane electrode was connected to ground through a series resistor which could be varied by factors of ten in steps from 100 ohms to 10^6 ohms. A coaxial cable carried the signal either through the preamplifier or directly to the synchroscope. The cable was terminated by its characteristic impedance for the study of fast pulses.

The electrically induced and the photomultiplier signals could be studied either separately or with one triggering the sweep of the synchroscope, while the other appeared on the screen. It was also possible to superimpose the two signals and study their time relation.

By means of a switch, a micro-ammeter could be set in series with the plane for measuring the current. Figure 1 shows a block diagram of the system.

In the early part of the work, i.e., for the study of burst pulses, the type 511D Tektronix sychroscope and type 121 Tektronix preamplifiers (rise time 0.06 microsecond) were used. However, for the study of faster pulses, a type 513D Tektronix (rise time 0.02 microsecond) became available. By direct application of pulses to the deflection plates of the cathode-ray tube, it was possible to resolve pulses with time intervals of the order of 10^{-8} second.

I. POSITIVE PRE-ONSET BURST PULSES

In the early days of corona studies, the burst pulses were first observed and named by Trichel[2] and in subsequent years were investigated by other workers in the field. As observed through a resistor, in series with the plane electrode (with increasing voltage), these pulses were the first type of disruptive discharge to appear before the onset of steady corona was reached in the positive point to plane geometry. They had a duration of a few milliseconds, and with the increasing voltage, they increased both in amplitude and duration. As observed on the oscilloscopes of the time, these pulses were not smooth, and the trace appeared with a number of ripples on top of it. Thus, due to the absence of fast synchroscopes and amplifiers, the detailed structure and, therefore, the mechanism of these pulses were not exactly known. However, it should be mentioned that the long duration of these pulses as compared with the short duration of streamers (10^{-6} second), together with the visual observation, helped to interpret these pulses as spreading laterally over the point in contrast to the streamers which propagate in the direction of axis toward the plane. It was furthermore indicated by L. B. Loeb, in interpreting Trichel's observations, that these pulses were analogous to the Geiger counter pulses observed in coaxial geometry at the lower

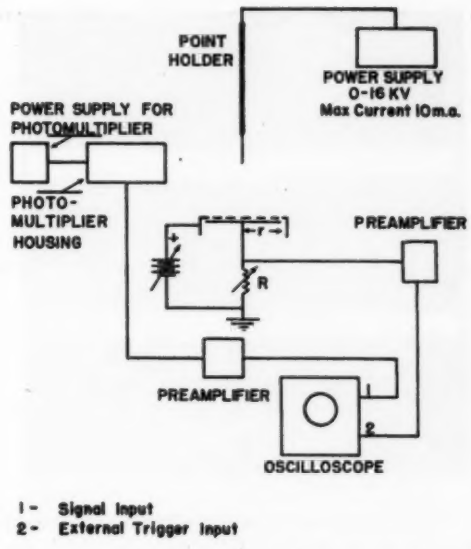


FIG. 1. Block diagram of the apparatus.

pressures. It was then believed by Loeb that they propagated over the point by photo-ionization in the gas and that they choked themselves off by space charge production as was later assumed also for fast Geiger counters.

The results presented here confirm the above belief and explain the long time duration of burst as well as their change with voltage and other variables. All the studies were made in room air at atmospheric pressure. Various sized points and gap distances were used; however, only a few representative oscillograms will be given here, and the results will be summarized in two tables.

1. The Electrically Induced Pulses

In order to compare these results with those of the other investigators,[4, 6] R (the series resistor) was set equal to 10^6 ohms. Then for a 1-mm point 4-cm gap in the presence of a source of ionizing γ radiation from thorionite ore, the following observations were made.

With the increase of voltage no discharges were observed until 7.3 kv was reached. At this point a single pulse of short duration (RC was made small for this observation), i.e., of the order of 10^{-6} second, appeared. At this threshold the first pulse was followed by pulses of the same nature, spaced at random a few milliseconds apart. As the voltage was increased to 7.32 kv, the initial pulse was followed by a *secondary pulse* of shorter duration and smaller height some 10 microseconds later. With further increase of the voltage the number of these secondary pulses increased, and the time interval between them became shorter. In all cases the sequence began with a large pulse followed by a number of secondary pulses. The amplitudes of both the initial pulse and the secondary pulses were also increased with increasing voltage. The time interval between the

FIG. 2. The electrically induced signals of burst pulses. 2.0-mm diameter point. 4-cm gap. $R=10^6$ ohms. $RC=10^{-4}$ sec. Sweep rate 100 μsec/scale division.

secondary pulses was random; however, a tendency for the intervals between secondary pulses to become shorter with increasing voltage was observed. Just before the onset of streamers (7.7 kv) the secondary pulses followed each other as closely as two microseconds apart, and on the average they were spaced 4 to 5 microseconds apart. The number of secondary pulses in a burst, therefore, varied from one up to a hundred or more, depending on the voltage. When a long RC time constant was used the initial pulse, plus the secondary pulses, gave the impression of a long single pulse with ripples on the top as observed by previous investigators.[4, 6, 7] This is illustrated in Figs. 2 and 3 and holds true for all the various cases studied. These pulses could be clearly resolved by decreasing RC. The primary and secondary pulses could be observed visually. However, their random nature prevented photographs being made. Thus the drawings shown in Fig. 4A, B, C, and D represent the visual sequence of the pulses above threshold. The interval between secondary pulses may be identified with Stever's[10] Geiger counter dead time. It should be added that since pulse duration was 10^{-7} sec the dead time $>2\mu$sec is nearly the same as the shortest time between pulses.

The effect of the radioactive source was prominent throughout the burst pulse region and especially at the threshold. In its absence the threshold was not at a definite value of voltage, and thus depended upon some accidental triggering. In the presence of the source the onset was definite within ± 20 volts in 7300 volts and always consisted of primary pulses some milliseconds or more apart.

At 7.7 kv where the streamers appeared, the burst pulses were by no means absent. As observed by Kip[4] and others, a streamer was generally followed by a burst pulse. However, in this case the initial large pulse was absent in the burst pulse sequence, and the burst pulse consisted of many small pulses following the streamer. With increasing voltage the streamers appeared more frequently, the duration of burst pulses increased, and finally, at 7.9 kv, the steady corona was achieved, shortly after pulse length became longer than the interval between pulses.

In order to measure the number of positive ions in the pulses, the time of crossing of positive ions from the pulses, and the clearing time (defined as the shortest time between two consecutive burst pulses), a screen with variable voltage source was set up similar to the one Kip used for the same purpose in the case of streamers. By partially uncovering the plane electrode, an inductive kick at the beginning of the trace was registered. This was followed by the collection of positive ions some time later, depending upon the crossing time of the ions to the plane. Knowing the sensitivity of the instruments and correcting for the number of ions removed by the screen, the total number of ions in the burst pulse was calculated. The screen was generally placed very close to the plane (2 to 4 mm). The auxiliary field was high so that to a good approximation, the time corresponding to the peak of ion current represented the time of crossing. Then the screen was completely removed, and these pulses were observed on a rather long time sweep. The first inductive kick was followed by the smoothly decreasing current pulse, which dropped rather sharply to zero after some few

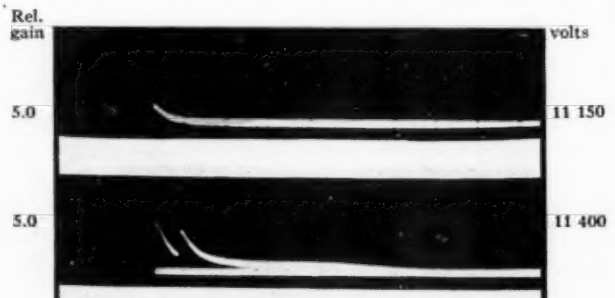


FIG. 3. The electrically induced signals of burst pulses. 2.0-mm diameter point. 4-cm gap. $R=10^6$ ohms. $RC=10^{-4}$ sec. Sweep rate 500 μsec/scale division.

[10] H. G. Stever, Phys. Rev. 61, 38 (1942).

milliseconds. Comparing the pulse without screen with that where the screen was present, it was found that the prolonged current was absent in the latter. This indicated that the long tail of the inductive pulse was due to the motion of positive ions across the gap in the low field region. Figures 3 and 5 illustrate this phenomenon rather nicely. The break in the long tail and the peak of ion current when the screen was present coincide (Fig. 5). The broadening of the break in the long tail was a diffusive effect due to a rather long time of formation of the burst pulse and nonuniformity of the field in the region of formation. The lowest trace on Fig. 5 was taken on a still longer time scale. Here the hump at the base of the sharp peak is the ion movement. The pulses at the extreme right indicate that the clearing time was somewhat shorter than the crossing time. In other words, the new burst pulse started when the space charge had crossed, say, only 80 percent of the gap. The number of ions measured by using the screen was the

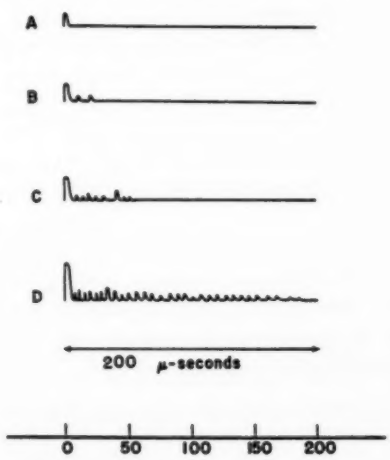


FIG. 4. Schematic representation of burst pulses as resolved and observed on the fast sweep rate of the oscilloscope. A at threshold; B, C, and D at various stages of the counting region.

total number of ions in the whole burst pulse (i.e., initial pulse and secondaries) and therefore subject to a large variation, depending on which part of the range of the burst pulse was under study.

A crude but useful way of checking the result would have been to estimate the number of ions in the individual initial and secondary pulses and then add these up to get the total number of ions. Since these pulses were very short (as will be shown later) and highly irregular, no direct way of measuring the number of ions was available. The pulse amplitudes were compared with those of streamer pulses which were of equally short duration, and the number of ions in which could be directly measured by the screen method. This comparison showed that the number of ions in the initial pulse varied from 0.01 to 0.02 of the streamer pulse, and that the secondary pulses had from 0.001 to 0.005 the number of ions in the streamers created under the same geometry. Thus, for the 1-mm point and 4-cm gap, the

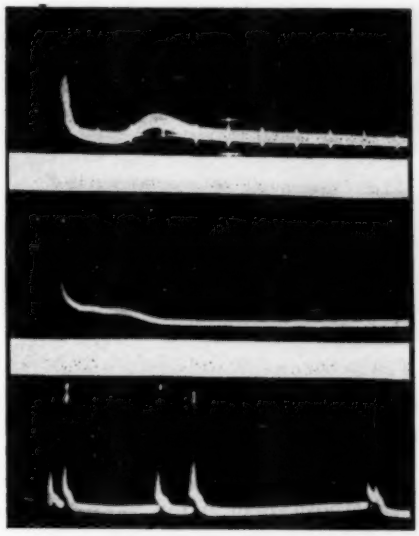


FIG. 5. The electrically induced signals of burst pulses. 2.5-mm diameter point. 4-cm gap. $R=10^6$ ohms. Upper: Using screen at a distance of 4 mm from the plane and at 200 volts. Sweep rate 1000 μsec/scale division. Middle: No screen. Sweep rate 1000 μsec/scale division. Lower: No screen. Sweep rate 5000 μsec/scale division.

number of ions in the initial pulse varied from 10^7 to 2×10^7 ions and in the secondary pulses varied from 10^6 up to 5×10^6 ions. When there were, say, one hundred secondary pulses in the burst pulse with average time intervals of five microseconds, a burst pulse of 500 microseconds duration gave some 5×10^8 ions. On the other hand, if there were only 10 secondary pulses the number of ions was 2×10^7.

This crude way of calculating the number of ions is in agreement with the method of direct measurement by means of the screen. Since the pulses were highly irregular no exact or suitable data of the number of ions *versus* voltage or point size were obtained, and the above figures merely indicate an estimate.

If we reduce R to 100 ohms ($RC\sim10^{-8}$ second) it is possible to study the initial pulse on the 0.1 microsecond/scale division sweep rate. This is shown together with the photon pulse observed by the photomultiplier in Fig. 6. The two will be compared and discussed later.

2. Pulses Observed by the Photomultiplier

Figure 6 shows the initial pulse of the burst pulse as observed by the photomultiplier. This is of some 1.2×10^{-7} second duration, which is the limit of the Tektronix type 511D and the pre-amplifier type 121. This indicates that the luminous phenomena associated with the initial pulse of the burst lasts less than 10^{-7} second. If we assume that excitation and emission of light parallel ionization, this duration indicates that ionization reaches its maximum and is choked off by local space charge in less than 10^{-7} second. Figure 6

FIG. 6. The initial pulse of the burst pulse. 2.5 mm diameter point. 4-cm gap. $R = 100$ ohms. $RC = 10^{-8}$ sec. Sweep rate 0.1 μsec/scale division. Upper: Photon pulse. Lower: Electrically induced pulse.

consists of superposition of many traces, thus indicating the reproducibility of the initial pulse.

Next, it was desirable to observe the secondary pulses and the whole burst pulse in order to confirm the picture obtained by electrical means and presented earlier. Since observation and photography of such short pulses on the long sweep time of the synchroscope was very difficult, the signal resistor of the photomultiplier was raised to 10^6 ohms. This increased the RC time constant of the circuit from 10^{-8} second to 10^{-4} second. Therefore, the pulses reached their maximum very rapidly but decayed according to the time constant of the circuit. The results obtained in this manner are illustrated in Fig. 7. It is seen that in principle photon and electrical observations Fig. 7 and 2 do not differ seriously.

At 7300 volts in Fig. 7 only a single primary pulse is seen. The other disturbances are "noise." At 7350 volts one single burst pulse is seen. The primary pulse cannot be distinguished but secondary pulses of variable intensity are clearly seen. At 7450 volts light from the primary and secondaries in the first burst pulse can be seen followed by a second burst pulse with primary pulse observed. It is to be noted that the electrical signals have a longer decay time than the photon signals, although the RC time constants for both cases were the same. This difference is due to the fact that the resistor in series with the plane electrode recorded the motion of positive ions in addition to the separation of charges, whereas the photomultiplier could not detect the motion of ions and merely recorded the time of active discharge.

After making the above comparison, it was desirable to observe the extent of the region of luminosity. This was done by applying a slit system to the photomultiplier. In this way observations could be confined to regions as close to the point surface as 1 mm. In fact, the point luminosity and the radius of curvature of 0.5 mm for the point prevented any closer study. It was observed that the luminosity was confined to within 1 mm of the surface. Visual observation confirmed this fact, since no extension of luminosity into the gap was noticed. In theory the fields for intensive ionization lie within 0.2 mm from the point. It is unfortunate that the photomultiplier observations were limited to 1 mm so that theory could not be confirmed.

3. Summary of Results

In Table I the voltage at the onset of burst pulses and streamers are tabulated for standard geometry, where the ratio of gap length to point diameter was kept equal to 80. For reference and comparison the values of Bandel[7] are quoted. It is seen that with this sensitive method of detection the threshold appears lower than previously observed. Room air at atmospheric pressure was used.

The data in Table II were taken for various sized points with 4-cm gap in room air at atmospheric pressure. The following notations are used in this table:

Range of burst pulses—this indicates the voltage range from the onset of burst pulses up to the onset of streamers.

Counter region—This indicates the voltage range from the onset of burst pulse up to the onset of steady corona.

Burst pulse duration—this is the total time taken for the spread and burning of the initial and secondary pulses in a burst pulse before extinction by space charge.

Time of crossing—this is the time taken for the positive ions to cross the gap.

Clearing time—this is the time taken for the space charge to move to a distance so that another burst pulse can start.

The times indicated above correspond to the notation introduced by Stever[10] for the geiger counters, using coaxial cylindrical geometry and lower pressures. He was able to measure the rise time, the combined time of

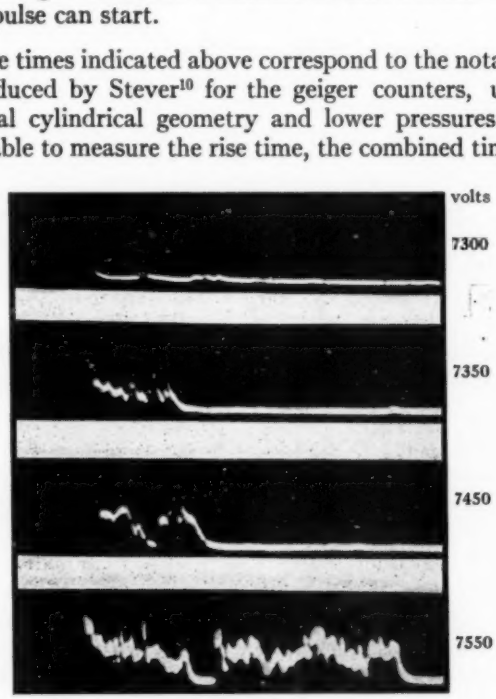


FIG. 7. The photon pulses of burst pulses. 1.0-mm diameter point. 4-cm gap. $RC = 10^{-4}$ sec. Sweep rate 1000 μsec/scale division

spread and burning, and the dead and clearing times of individual pulses. For a small point, it was not possible to resolve the rise time, the time of spread, and the burning time with the present equipment. However, the dead time, defined as the time between two successive individual pulses of a burst pulse, as well as the clearing time, could be measured. The dead time was a variable quantity and very irregular; however, near the onset of streamers it could become as short as 2 microseconds.

It should be mentioned that the times given here are the most probable ones, and due to the random occurrence of these pulses, shorter and longer burst pulse durations have been observed at a given voltage.

4. Interpretation of the Data

From the above results and the illustrative oscillograms, the mechanism of burst pulses is clarified in the following manner. As soon as the field conditions are proper, an electron appearing in the gap, forms an avalanche according to the $\exp(\int \alpha dx)$ law. Associated with the avalanche formation there is excitation as well as ionization. The mixed gas called air, being a good photoabsorbent of its own photons, and being ionized by its own photons, helps the initial pulse to spread over the point by photoionization in the gas in less than 10^{-7} second. To confirm this picture, reference is made to Fig. 6. The photon pulse indicates only the short time of ionization and spread, whereas the electrical pulse shows the short time of ionization and spread followed by the motion of positive ions in the high field region after the ionization had ceased. Once the initial pulse has spread over the point, it is choked off by its own space charge. Owing to the nonuniformity of the field over the point, a region of the point is partially cleared before the whole point is cleared, and a second pulse, triggered by the shower of photons, is permitted to form and spread. Because of the presence of the space charge of the initial pulse, the second pulse is not permitted to spread very far and, being quickly quenched, has a smaller amplitude. This process is repeated until enough space charge is accumulated in the relatively lower field regions of the gap to choke the whole burst. No other pulse can take place until this space charge is cleared. When the space charge has moved far enough across the gap to the cathode, another burst consisting of a primary with its group of secondaries is permitted to take place. Thus, depending upon the field conditions, there may be one, two, or hundreds of pulses in a burst. This phenomenon is similar, but with differences, in its action to that of the *self-quenching Geiger counters*, since it involves dead and clearing times.

TABLE I.

Point diameter in mm	0.25	0.5	1.0
Bandel's values of onset of burst pulse in KV	3.49	5.06	7.67
Amin's values of onset of burst pulse in KV	3.45	5.00	7.50
Bandel's value of onset of streamers in KV	···	5.26	7.98
Amin's value of onset of streamers in KV	3.55	5.26	7.90

TABLE II.

Point in diam in mm	0.50	1.00	2.00	2.50
Range of burst pulse in KV	5.00–5.26	7.30–7.70	11.15–11.73	12.85–13.60
Range as percentage of onset of burst pulse voltage	5.20	5.48	5.21	5.84
Geiger counter region in KV	5.00–5.28	7.30–7.90	11.15–11.87	12.85–13.90
Burst pulse duration at the end of the range in msec	0.4–0.5	0.5–0.7	0.8–1.0	1.0–1.2
Time of crossing at the end of the range in msec	9.5	4.5	2.2	1.5
Minimum clearing time in msec	6.0	2.5	1.5	1.0

At this point it is essential to point out that choking action by space charge in strongly asymmetrical gaps involves two mechanisms: (1) a quick, short-lived quenching with rapid clearing caused by accumulation near the point; and (2) a slow quenching with long clearing time caused by mass accumulation of space charge in the lower field regions of the gap.

In the study of Geiger counters and the analysis of counting plateau, the study has been generally made in connection with a recording circuit, responding to a certain pulse amplitude. Thus, a study of what happens as potential on a self-quenching corona counter is raised, has never been delineated in terms of physical phenomena occurring. While the geometry of point-to-plane gap is not adapted to accurate counting, the time scale analysis with increasing potential made here provides a clear physical picture of the mechanism of corona counters.

With a positive point and negative plane electrons created by γ rays from the thorionite ore are driven to the point. When the potential reaches the value setting the threshold for the burst pulse propagation, as given by Loeb,[11] single pulses of the order of 10^7 ions occur at time intervals of several milliseconds. These pulses are such that the space charge created near the point by the 10^7 ions (produced in less than 10^{-7} second) is able to quench the discharge, and the discharge threshold is so sensitive to space charge that no new count can occur until the clearing time of milliseconds or so has elapsed. It must also be noted that the secondary pulses with a weak γ-ray source are triggered not by the external counting source but by photoelectrons produced from the first pulse in the body of the gas. In other words, as potential reduces the effectiveness of the space charge and a so-called counting plateau is approached, secondary pulses in an increasing number appear following the initial pulse. The secondary pulses are inherently the be-

[11] L. B. Loeb, J. Appl. Phys. **19**, 882 (1948).

ginning of the phenomenon known to Geiger counter workers as self-counting.

It must be noted at this juncture that the peculiar point-to-plane geometry and the use of air at 760 mm leads to local triggering of counts after the dead time which does not occur in well made organically quenched coaxial cylindrical counters at lower pressures and potentials.

Thus, such phenomena are not as clearly observed in Geiger counters where the time of spread, etc., down the wire gives longer dead times and larger secondary pulses, and where the gases used are such as to limit photoelectric ionization much closer to the wire than is the case for air.

Eventually, as potential is further raised, the number of burst pulses, the number of secondary pulses per burst, and thus the duration of each burst pulse is increased, and the clearing time is decreased. With photoelectrons being produced further and further out in the gap with increasing duration of bursts, and creation of O_2^- ions from these the gap is filled with self-created triggering electrons. Accordingly many more individual burst pulses are observed than can be accounted for by triggering electrons from the external radioactive source. Thus, self-counting sets in with a very small plateau in this geometry.

Apparently also when the field about the point at a potential V_{str} achieves ionizing values over distances that are great enough (large points, small gaps) with water vapor present the intensity of photoionization in the burst pulse near the tip leads to the building outward of a positive space charge boss. The field distortion created by this with the help of photoionization initiates a streamer. Then, as noted before, a mixture of burst pulses and streamers triggering burst pulses follow. The discussion of the appearance of streamers must, however, be postponed to a later section of the report.

With increasing potentials streamer propagation is favored, and streamers become more frequent for some range of values of V above V_{str}. However, observation never shows a streamer occurring *during* a burst pulse. It is observed that a streamer initiates during the growth of the primary pulse of a burst, i.e., within 2×10^{-8} sec. It does not occur once the burst pulse space charge is about the point. A burst pulse of secondaries may follow a streamer acting as a primary pulse. Thus, while burst pulses burn and until they clear, streamers cannot form.

Since burst pulses can generally develop more readily than the streamers (i.e., burst pulses are usually several times more probable than streamers) above the streamer threshold with increasing potential, the lengthening burst pulse with shorter clearing time leaves little chance for streamer formation. Therefore, when the applied potential reaches this point the burst pulses are almost continuous with little interrruption. Streamers disappear, and the burst pulses shortly merge into a self-sustaining continuous but slightly fluctuating corona.

The writer wishes to express his thanks to Professor L. B. Loeb at whose suggestion this study was undertaken, for his kind advice and encouragement.

JOURNAL OF APPLIED PHYSICS VOLUME 25, NUMBER 2 FEBRUARY, 1954

Parallel-Connected Magnetic Amplifiers

S. H. CHOW
Burroughs Corporation, Philadelphia, Pennsylvania
(Received June 10, 1953)

A complete analysis of parallel-connected magnetic amplifiers, based on a single-valued B-H relationship, is given. Three different types of loads: resistive, inductive, and capacitive are considered. Transient response is studied. Analysis of feedback amplifiers clearly indicates the presence of the "jump" phenomenon.

TABLE OF SYMBOLS

A	Area of magnetic flux path
B	Magnetic flux density
B_0	Average flux density
B_t	Transient input flux density
b	Magnetic flux density (instantaneous value)
$\dot{b}$	db/dt
C	Capacitance
E	Voltage (amplitude)
E_i	Input voltage
e	Voltage (instantaneous value)
F	Magnetomotive force
H	Magnetic field intensity
h	Magnetic field intensity (instantaneous value)
h_A	Magnetic field intensity in core A
h_B	Magnetic field intensity in core B
I	Electric current
i	Electric current (instantaneous value)
i_l	Load current
i_p	Current through carrier winding
k	Coefficient in B-H relationship $H=\Sigma k_n B^n$
L_c	Inductance of carrier winding
L_l	Load inductance
$\mathbf{L}$	Laplace operator $\mathbf{L}[f(t)]=\bar{f}(p)$
$\mathbf{L}^{-1}$	Inverse Laplace operator $\mathbf{L}^{-1}[\bar{f}(p)]=f(t)$
l	Length of magnetic flux path
N	Number of turns

N_f Number of turns of feedback winding
R Resistance
R_i Input resistance
R_l Load resistance
μ Permeability

INTRODUCTION

THE ideal B-H characteristic curve for a parallel-connected magnetic amplifier is shown in Fig. 1(a). In order to make a mathematical analysis possible the B-H relationship will be represented by

$$H=k_1B+k_3B^3, \tag{1}$$

shown in Fig. 1(b). The treatment of steady state response follows the method used by Pipes,[1] but leads to entirely different results. The reason for this has been pointed out by Kirschbaum.[2] Transient response is studied by the same method, considering the input to be a step function. In treating feedback amplifiers, in which current at even harmonic frequencies is a significant factor, the second harmonic component as well as the dc and fundamental components are considered. In this paper carrier and input windings are considered to have the same number of turns. Ordinary transformer theory applies when they are different.

AMPLIFIERS WITH RESISTIVE LOAD

A. Steady State

The circuit is shown in Fig. 2(a). When the load resistance R_l is low ($R_l \approx 0$), the voltage across the carrier coils is the same as the applied carrier voltage. The input current $I_i=E_i/R_i$ produces the flux density B_0 which may be readily found from the B-H curve. The carrier voltage $E\sin\omega t$ produces the swing in flux density $B=E/NA\omega$.

The instantaneous flux densities in core A and core B are

$$b_A=B_0-B\cos\omega t, \tag{2a}$$

$$b_B=B_0+B\cos\omega t, \tag{2b}$$

and the corresponding magnetic field intensities are

$$h_A=k_1b_A+k_3b_A^3, \tag{3a}$$

$$h_B=k_1b_B+k_3b_B^3. \tag{3b}$$

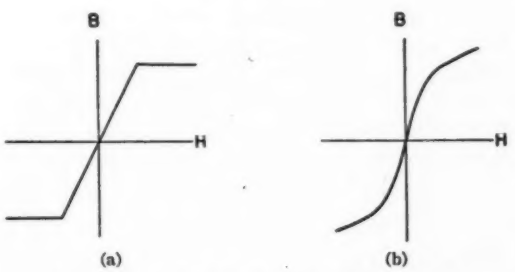


FIG. 1. B-H characteristics.

[1] L. A. Pipes, J. Appl. Phys. **23**, 625 (1952).
[2] H. S. Kirschbaum, J. Appl. Phys. **23**, 1278 (1952).

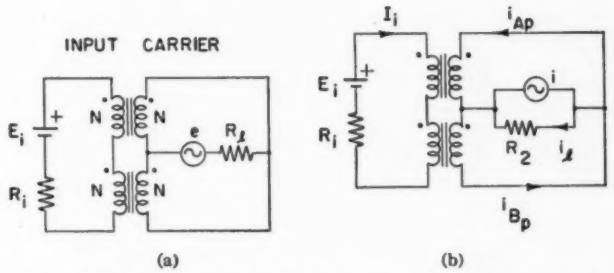


FIG. 2(a). Parallel-connected amplifier. (b) Parallel-connected amplifier: voltage output.

The load current is therefore

$$i_l=(l/N)(h_A-h_B)$$
$$=-\frac{2k_1l}{N}B\cos\omega t-\frac{2k_3l}{N}(3B_0^2B\cos\omega t+B^3\cos^3\omega t). \tag{4}$$

The load current contains fundamental and third harmonic components. For other B-H characteristics the load current contains different odd harmonic components. The currents in the carrier windings contain circulating dc and second-harmonic components in addition to the odd harmonic components of the load current. The useful component of the load current, i.e., the component that changes with input, is the term $[-(6k_3lB_0^2B)/N]\cos\omega t$ in Eq. (4). Figure 3 shows the relation between i_l and B_0 and the corresponding relation between i_l and I_i.

When the load resistance is not negligible, the voltage across the carrier windings is less than the applied voltage. The flux density is no longer sinusoidal. Equations (2) and (4) become

$$b_A=B_0+b(t), \tag{5a}$$

$$b_B=B_0-b(t), \tag{5b}$$

and

$$i_l=(l/N)(h_A-h_B)$$
$$=\frac{2k_1l}{N}b+\frac{2k_3l}{N}(3B_0^2b+b^3), \tag{6}$$

where $b(t)$ is a periodic function to be determined.

Apply Kirchoff's law to the carrier circuit,

$$E\sin\omega t=NA\dot{b}+R_li_l, \tag{7}$$

or

$$\dot{b}+\frac{R_ll}{N^2A}(2k_1+6k_3B_0^2)b=\frac{E\sin\omega t}{NA}-\frac{2k_3l}{N^2A}b^3. \tag{8}$$

Put

$$[(R_ll)/(N^2A)](2k_1+6k_3B_0^2)=m$$

and

$$(2k_3R_ll)/N^2A)=a.$$

Equation (8) then becomes

$$\dot{b}+mb=\frac{E\sin\omega t}{NA}-ab^3. \tag{9}$$

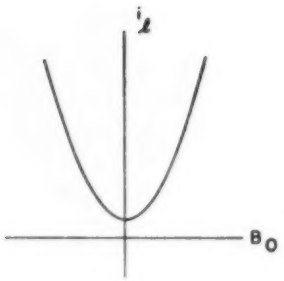
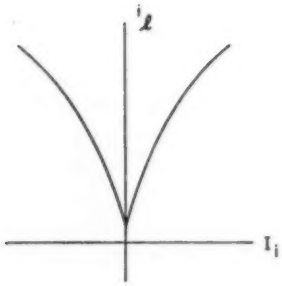


FIG. 3. Input-output characteristics.



If the term ab^3 is neglected, Eq. (9) represents a linear LR circuit, L being the inductance of the two carrier windings connected in parallel and biased at B_0 and R being the load resistance. This can be seen from the following relations:

$$H=k_1B+k_3B^3,$$
$$1/\mu=(dH/dB)=k_1+3k_3B^2.$$

Therefore,

$$m=R_l\Big/\left(\frac{1}{2}\frac{\mu N^2A}{l}\right)=2R_l/L_c.$$

Equation (9) can be solved in the following manner. Take the Laplace transform of Eq. (9), and denote $\mathbf{L}[b(t)]$ by $\bar{b}(p)$, where $\mathbf{L}$ is the Laplace operator. It is found that

$$\bar{b}(p)=\frac{E\omega}{NA(p+m)(p^2+\omega^2)}-a\frac{\mathbf{L}[b^3(t)]}{p+m}. \tag{10}$$

Equation (10) corresponds to the case in which a switch in series with the carrier source is closed at $t=0$. The resulting transient response is entirely different from the transient response produced by an input step function. The latter is the transient response of interest.

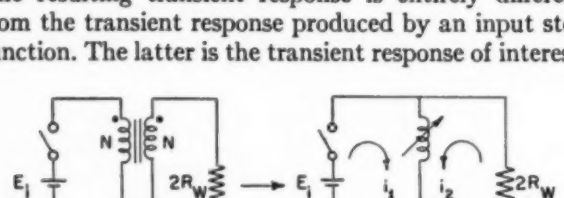


FIG. 4. Equivalent circuit for parallel-connected amplifier: transient response.

In the present case the Laplace transform is used only as a convenient means for obtaining the steady state response of Eq. (9). Its use will lead to correct results since the steady state response of this particular nonlinear equation does not depend on initial conditions. No physical significance will be attached to the transient portion of the solution.

Take the inverse transform of Eq. (10). Then

$$b(t)=\frac{E\omega}{NA}\left[\frac{\epsilon^{-mt}}{m^2+\omega^2}+\frac{\sin(\omega t-\theta)}{\omega(m^2+\omega^2)^{\frac{1}{2}}}-a\int_0^t\epsilon^{-m(t-u)}b^3(u)du\right], \tag{11}$$

where

$$\theta=\tan^{-1}(\omega/m).$$

This integral equation can be solved by the method of successive approximations, with the first approximation taken as

$$b^{(1)}(t)=\frac{E}{NA}\frac{\sin(\omega t-\theta)}{(m^2+\omega^2)^{\frac{1}{2}}} \tag{12}$$

with transient neglected.

The second approximation is

$$b^{(2)}(t)=b^{(1)}(t)-a\int_0^t\epsilon^{-m(t-u)}[b^{(1)}(u)]^3du, \tag{13}$$

etc. The load current can be found from Eq. (6) and contains only odd harmonic components.

When a high-voltage output is desired, a constant amplitude ac current I may be sent through the reactors and a high load resistance R_l connected in parallel with the carrier windings, as shown in Fig. 2(b). The circuit equation for this case

$$\dot{b}+\frac{R_l}{N^2A}(2k_1+6k_3B_0^2)b=\frac{R_lI\sin\omega t}{NA}-\frac{2R_lk_3}{N^2A}b^3 \tag{14}$$

is necessarily the same as Eq. (8), with E changed to R_lI. The load current is

$$i_l=(NA\dot{b})/R_l. \tag{15}$$

B. Transient Response

In parallel-connected amplifiers, the response time is slow because any change in input current is short-circuited by the two carrier windings connected in parallel. The equivalent circuit may be simplified to that shown in Fig. 4 because $R_l\gg R_w$ where R_w is the winding resistance.

When a control voltage E_i in the form of a step function is applied to the input with no carrier voltage applied to the carrier winding, the equations for transient response are (refer to Fig. (4) for branch current notations)

$$E_i=R_ii_1+2NA\dot{B}_t, \tag{16}$$
$$0=2R_wi_2+2NA\dot{B}_t, \tag{17}$$

where B_t is the transient input flux density. On combining Eqs. (1), (16), and (17),

$$\dot{B}_t+\frac{l}{N^2A}(k_1B_t+k_3B_t^3)\frac{R_iR_w}{R_i+2R_w}=\frac{E_i}{NA}\frac{R_w}{R_i+2R_w}, \quad (18)$$

or

$$\dot{B}_t+\alpha B_t+\beta B_t^3=\gamma. \quad (19)$$

The Laplace transform of Eq. (19) gives

$$\bar{B}_t(p)=\frac{\gamma}{p(p+\alpha)}-\frac{\beta\mathbf{L}[B_t^3]}{p+\alpha}. \quad (20)$$

The inverse transform gives

$$B_t=\frac{\gamma}{\alpha}(1-\epsilon^{-\alpha t})-\beta\int_0^t\epsilon^{-\alpha(t-u)}[B_t^3(u)]du. \quad (21)$$

Using the method of successive approximations, with the first approximation taken as

$$B_t^{(1)}=(\gamma/\alpha)(1-\epsilon^{-\alpha t}), \quad (22)$$

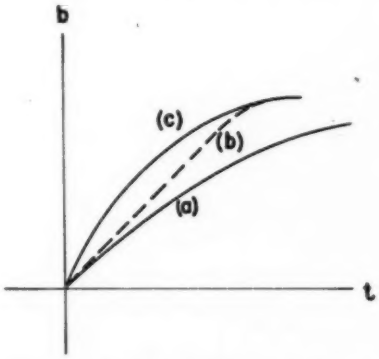


FIG. 5. Transient flux densities in linear and nonlinear LR circuits: (a) $=L=2L_{c(B=0)}$, (b) = nonlinear L, and (c) $=L=2L_{c(B=B_0)}$.

the second approximation is

$$B_t^{(2)}=B_t^{(1)}-\beta\int_0^t\epsilon^{-\alpha(t-u)}[B_t^{(1)}(u)]^3du, \quad (23)$$

etc.

The rise of input flux in a nonlinear LR circuit is not different in nature from that in a linear LR circuit. In the present circuit, when the input flux is first rising, the permeability of the two cores is $[1/(dH/dB)]_{B=0}=1/k_1$. When it reaches B_0 the permeability is $1/(k_1+3k_3B_0^2)$. If the transient flux densities for linear circuits with $L=2L_{c(B=0)}$ and $L=2L_{c(B=B_0)}$ are plotted, the transient flux density for the nonlinear circuit (Eq. 18) lies between the two curves (Fig. 5).

The above result represents the case when the carrier power is not turned on. The rise of flux density and the corresponding differential permeability $1/(k_1+3k_3B^2)$ are shown in Fig. 6(a). With carrier power on, the total flux densities and the corresponding differential permeabilities of core A and core B are shown in Figs. 6(b) and 6(c). Since in transient studies the two windings

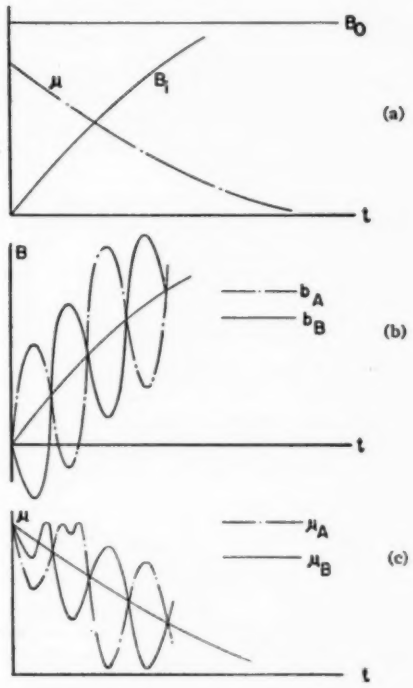


FIG. 6(a) Input flux density and corresponding permeability. (b) Total flux densities in core A and core B. (c) Corresponding permeabilities of core A and core B.

are effectively in series, the inductance of the two coils in series is very close to the value when ac flux is neglected.

AMPLIFIERS WITH INDUCTIVE LOAD

The circuit is shown in Fig. 7. Apply Kirchoff's law to the carrier circuit,

$$E\sin\omega t=NA\dot{b}+L_l\frac{di_l}{dt}, \quad (24)$$

where L_l is the load inductance. The load current i_l is again given by Eq. (6). Substitute Eq. (6) into Eq. (24) and integrate the resultant equation,

$$b^3+\left(\frac{2L_l+L_c}{L_lL_c}\right)\frac{N^2A}{2k_3l}b+\frac{NE\cos\omega t}{2L_l\omega k_3}=0, \quad (25)$$

where L_c is the carrier winding inductance at the bias $B=B_0$.

Equation (25) is of the form

$$b^3+qb+r=0. \quad (26)$$

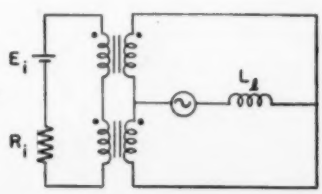


FIG. 7. Parallel-connected amplifier with inductive load.

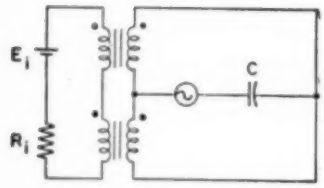


FIG. 8. Parallel-connected amplifier with capacitive load.

Since q is positive, the equation has only one real root

$$b=(2/\sqrt{3})q^{\frac{1}{2}}\sinh(\phi/3), \tag{27}$$

where

$$\sinh\phi=(3/q)^{\frac{1}{2}}(r/2).$$

The load current is then given by Eq. (6).

AMPLIFIERS WITH CAPACITIVE LOAD

The circuit is shown in Fig. 8. The circuit equation is

$$E\sin\omega t=NA\dot{b}+\frac{1}{C}\int(i_{AP}-i_{BP})dt,$$

or

$$\ddot{b}+\frac{l}{N^2AC}(2k_1+6k_3B_0^2)b=\frac{\omega E\cos\omega t}{NA}-\frac{2k_3l}{N^2AC}b^3. \tag{28}$$

Put $[l/(N^2AC)](2k_1+6k_3B_0^2)=m^2$. The Laplace transform of Eq. (28) reads

$$(p^2+m^2)\bar{b}(p)-px_0-x_1=\frac{\omega E}{NA}\frac{p}{p^2+\omega^2}-\frac{2k_3l}{N^2AC}\mathbf{L}[b^3], \tag{29}$$

where x_0 and x_1 represent initial conditions.

Take the inverse transform of Eq. (29). It is found that

$$b(t)=\frac{\omega E}{NA}\frac{1}{m^2-\omega^2}[\cos\omega t-\cos mt]+x_0\cos mt+\frac{x_1}{m}\sin mt-\frac{2k_3l}{N^2AC}\mathbf{L}^{-1}\left[\frac{\mathbf{L}[b^3]}{p^2+m^2}\right]. \tag{30}$$

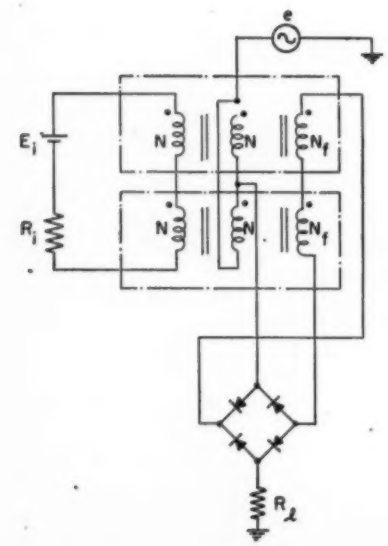


FIG. 9. Feedback amplifier.

In Eq. (30) m represents the angular velocity of self-oscillation. In this circuit no resistance has been included. A small resistance will damp out the self-oscillation. Therefore,

$$b(t)=\frac{\omega E}{NA}\frac{\cos\omega t}{m^2-\omega^2}-\frac{2k_3l}{N^2AC}\mathbf{L}^{-1}\left[\frac{\mathbf{L}[b^3]}{p^2+m^2}\right]. \tag{31}$$

Proceeding as before, the first approximation to b may be taken as

$$b^{(1)}(t)=\frac{\omega E}{NA}\frac{\cos\omega t}{m^2-\omega^2}. \tag{32}$$

Since

$$\frac{1}{p^2+m^2}=\frac{1}{2jm}\left(\frac{1}{p-jm}-\frac{1}{p+jm}\right),$$

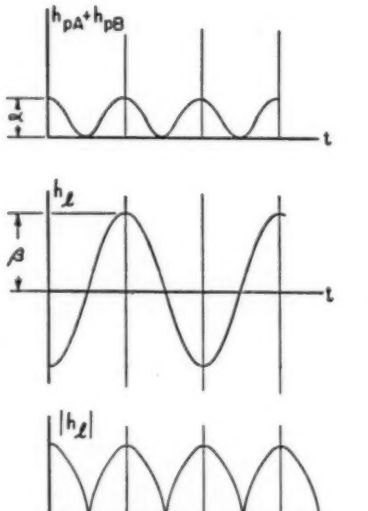


FIG. 10. Current waveforms in feedback amplifiers.

the second approximation is

$$b^{(2)}(t)=b^{(1)}(t)-\frac{2k_3l}{N^2ACm}\int_0^t\sin m(t-u)\times[b^{(1)}(u)]^3du, \tag{33}$$

etc.

Series tuning in a nonlinear circuit increases the load current in the same manner as in a linear circuit, and therefore may be used to increase the gain over a limited range of input current. It should be mentioned, however, that Eq. (28) permits sustained subharmonic oscillations when k_3 is high. In such a case, the method of analysis used above will not give correct results. Also a "jump" phenomenon exists within a certain frequency range. If this is the case, detuning becomes necessary for successful amplifier operation.

FEEDBACK AMPLIFIERS

Feedback is commonly used to increase the sensitivity of a magnetic amplifier. It is accomplished by passing the rectified output current through additional feedback windings. A commonly used feedback

amplifier is shown in Fig. 9. In the following analysis, the diodes are considered to be perfect.

It has been shown that in parallel-connected amplifiers there are no even harmonic components in the load current. When the load resistance is low, the voltage across the carrier is purely sinusoidal and the flux densities in the two cores are given by Eq. (2), where B_0 is now the resultant average flux density, including both input and feedback components.

The magnetomotive forces in core A and core B are

$$F_A=N(I_i+i_{AP})+N_f|i_l|, \tag{34a}$$

$$F_B=N(I_i+i_{BP})+N_f|i_l|, \tag{34b}$$

where i_{AP} and i_{BP} are carrier winding currents.

Denote Ni/l by h. The magnetic field intensities are

$$h_A=H_i+h_{AP}+(N_f/N)|h_l|, \tag{35a}$$

$$h_B=H_i+h_{BP}+(N_f/N)|h_l|, \tag{35b}$$

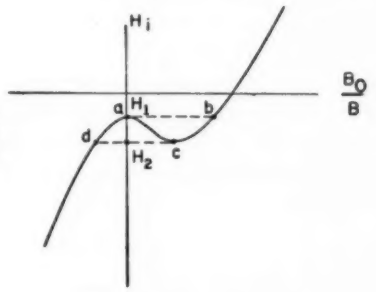


FIG. 11. Input current *vs* resultant dc flux density in feedback amplifiers.

in which h_{AP} and h_{BP} contain both even and odd harmonics. To separate them, express h_{AP} by

$$h_{AP}=\tfrac{1}{2}(h_{AP}+h_{BP})+\tfrac{1}{2}(h_{AP}-h_{BP})$$
$$=\tfrac{1}{2}(h_{AP}+h_{BP})+\tfrac{1}{2}h_l. \tag{36}$$

The waveforms of $(h_{AP}+h_{BP})$, h_l, and $|h_l|$ are shown in Fig. 10 when only dc, fundamental, and second harmonic currents are considered. They have the form

$$h_{AP}+h_{BP}=\frac{\alpha}{2}(1+\cos 2\omega t), \tag{37a}$$

$$h_l=-\beta\cos\omega t, \tag{37b}$$

$$|h_l|=\beta\left(\frac{2}{\pi}+\frac{2}{3\pi}\cos 2\omega t\right), \tag{37c}$$

and h_A has the form

$$h_A=\left(H_i+\frac{\alpha}{4}+\frac{2N_f\beta}{N\pi}\right)-\frac{\beta}{2}\cos\omega t$$
$$+\left(\frac{N_f}{N}\frac{2\beta}{3\pi}+\frac{\alpha}{4}\right)\cos 2\omega t. \tag{38}$$

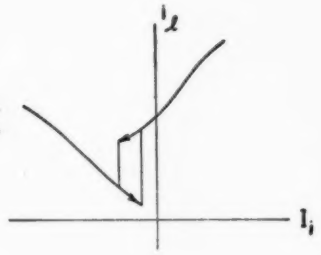


FIG. 12. Input-output characteristics in feedback amplifiers.

From the B-H relationship [Eq. (1)]

$$h_A=k_1(B_0-B\cos\omega t)+k_3(B_0-B\cos\omega t)^3. \tag{39}$$

Equate the dc, fundamental, and second harmonic components in Eqs. (38) and (39),

$$H_i+\frac{\alpha}{4}+\frac{2N_f\beta}{N\pi}=k_1B_0+k_3\left(B_0^3+\frac{3B_0B^2}{2}\right), \tag{40a}$$

$$\beta/2=k_1B+k_3(3B_0^2B+\tfrac{3}{4}B^3), \tag{40b}$$

$$\frac{\alpha}{4}+\frac{2N_f\beta}{3N\pi}=k_3(\tfrac{3}{2}B_0B^2). \tag{40c}$$

On eliminating α and β,

$$\left(\frac{B_0}{B}\right)^3-\frac{8}{\pi}\left(\frac{B_0}{B}\right)^2+\frac{k_1}{k_3B^2}\left(\frac{B_0}{B}\right)$$
$$-\left(\frac{H_i}{k_3B^3}+\frac{2}{\pi}\frac{N_f}{N}+\frac{8}{3\pi}\frac{k_1}{k_3B^2}\frac{N_f}{N}\right)=0. \tag{41}$$

Equation (41) governs the stability of the amplifier. For an amplifier with fixed applied voltage (or B) and feedback ratio N_f/N, B_0/B is triple-valued in a certain range of H_i, as shown in Fig. 11. This curve shows that when H_i is increased from a low value to a value slightly beyond H_1, the average flux density jumps from point a to point b. And when H_i is decreased from a high value to a value slightly below H_2, the average flux density jumps from point c to point d. Since the useful load current is proportional to B_0^2 for the B-H relationship [Eq. (1)], the output current-input current curve is that shown in Fig. 12.

When the load resistance is not low, any change in load current will be opposed by the resistive drop across the resistance. Therefore the effect of feedback is partly suppressed.

Current Noise in Thermistor Bolometer Flakes*

JAMES J. BROPHY
Armour Research Foundation of Illinois Institute of Technology, Chicago, Illinois

(Received June 29, 1953)

The characteristics of the additional electrical noise generated by iron oxide thermistor bolometer flakes due to direct current passing through the units is investigated. It is found that the noise is similar to the current noise observed in other semiconductors. The current noise power density is observed to depend inversely upon frequency, to be approximately proportional to the direct current, and to be independent of temperature. It is indicated that the major source of the noise is at the electrodes which make electrical contact to the semiconductor. Electrodes produced by evaporated gold applied through a potential difference are found to yield relatively low noise flakes.

I. INTRODUCTION

THE large negative temperature coefficient of resistance characteristic of the class of semiconducting materials known as thermistors makes it possible to employ these materials as sensitive detectors of infrared radiation. A recent paper[1] indicates the usefulness of thermistor bolometer flakes in this application.

The ultimate sensitivity of thermistor bolometers is limited by the generation of Johnson or thermal noise in the units. It is found, however, that in practice the sensitivity may be considerably less than ideal because of the generation of additional electrical noise which appears when a direct current is passed through the flakes. The current is required in order to determine the change in resistance of the bolometers when subjected to radiation.

The appearance of additional electrical noise with the passage of a direct current through semiconductors is well established.[2–4] It is known that it may be associated with poor electrical contacts between semiconductors,[2] or between a semiconductor and a metallic conductor.[3] It has recently been shown that current noise exists in semiconductors even in the case of "noiseless" electrodes.[4] The phenomenon of current noise has yet to receive a completely satisfactory, comprehensive, theoretical treatment.

The current noise problem in bolometer flakes of standard thermistor materials (mixtures of oxides of manganese, nickel, and cobalt) appears to be well in hand.[1] It is found for these materials that a major portion of the noise can be attributed to the method of applying electrodes to the flake. It does not appear that a detailed investigation of the current noise generated by these units has been reported, however,

In this paper the results are given of a study of the current noise produced by thermistor bolometers constructed from an iron oxide thermistor mixture.[5] The purpose of the investigation was to examine the additional noise arising from the passage of direct current through the bolometer and to attempt to distinguish between internal noise sources and those due to electrode properties.

II. APPARATUS

To investigate the current noise generated by the bolometer flakes a low-noise, high-gain, variable bandwidth amplifying chain was set up which, except for the output measuring unit and the input circuit, consists of standard commercial equipment. The system is capable of a useful voltage gain of 10^7 over a frequency range of one cylce per second to 20 kilocycles, has variable bandwidth and variable center frequency, and yields a true rms measure of the input voltage, independent of wave form.

A block diagram of the circuit is shown in Fig. 1 together with the detailed diagram of the input circuit. This circuit was constructed separately in a well-shielded metal box and connected to the preamplifier by means of shielded cable. The 400 000-ohm wire wound resistor and 4-mfd oil capacitor constitute a resistance-capacity filter to eliminate battery noise from the input of the amplifier. The variable ballast resistor is a number of non-inductive wire wound units selected by a tap switch to match the sample resistance. In the figure, C_i represents the stray wiring shunt capacitance as well as the input capacitance of the preamplifier.

The preamplifiers used are Tektronix Type 122; the band pass filter is a Krohn-hite Model 330-A and the oscilloscope is a du Mont 304-A. It was found convenient to use the vertical amplifier of the oscilloscope as part of the amplifying chain to achieve the aquired amplification. The rms vacuum tube voltmeter, which was connected to

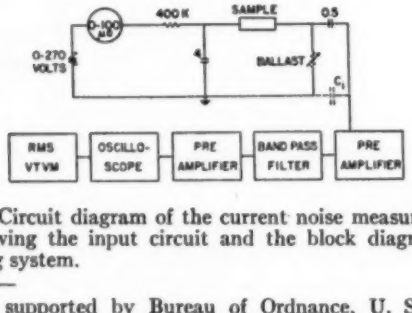


FIG. 1. Circuit diagram of the current noise measuring equipment showing the input circuit and the block diagram of the amplifying system.

* Work supported by Bureau of Ordnance, U. S. Navy.

[1] E. M. Wormser, J. Opt. Soc. Am. 43, 15 (1953).

[2] C. J. Christensen and G. L. Pearson, Bell System Tech. J. 15, 197 (1936).

[3] P. H. Miller, Proc. Inst. Radio Engrs. 35, 252 (1947).

[4] H. C. Montgomery, Bell System Tech. J. 31, 950 (1952).

[5] Supplied by Horizons, Inc. of Clevelend, Ohio.

the output terminals of the scope, consisted of a simple 6J6-cathode follower feeding a standard, one milliampere thermocouple.

The linearity and the frequency response of the system was checked by feeding to the sample terminals sine wave signals obtained from a signal generator and a General Radio Microvolter. The shape of all pass bands used in the study was plotted and the mid-band gain and equivalent bandwidth[6] determined by graphical integration. The stray capacitance C_i was measured.

The open-circuit, mean-square current-noise voltage per unit band width was calculated from the expression

$$v_c^2/B=[4+R^2\omega^2C_i^2/G^2B_{eq}](V^2-V_0^2), \qquad (1)$$

where R is the dynamic sample resistance, ω is the center frequency of the pass band, C_i is the stray shunt capacitance, G is the mid-band gain, B_{eq} is the equivalent

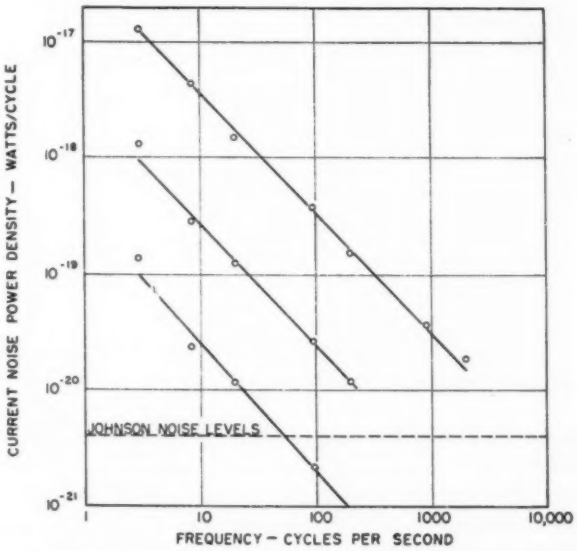


FIG. 2. The current noise power density spectrum of three thermistor bolometer flakes for a direct current of 60 microamperes. The Johnson noise level is calculated from Nyquist's equation, $P_J=kT$ watts/cycle, where k is Boltzman's constant and T is the absolute temperature.

bandwidth, V is the rms output observed with current passing through the sample, and V_0 is the rms output for no current. This expression is for the case of the ballast resistance equal to the dynamic flake resistance and assumes the statistical independence of current and thermal noise. The output V_0 is due to thermal plus amplifier noise. Equation (1) is obtained by straightforward circuit analysis of Fig. 1.

To check the operation of the system the Johnson noise levels of several resistances were measured. It was found that the results agreed both in magnitude and frequency dependence with Nyquist's equation.[7,8] Finally the current noise of a wirewound resistor and a carbon

[6] F. E. Terman and J. M. Pettit, *Electrical Measurements* (McGraw-Hill Book Company, Inc., New York, 1952), p. 359.
[7] J. B. Johnson, Phys. Rev. 32, 97 (1928).
[8] H. Nyquist, Phys. Rev. 32, 110 (1928).

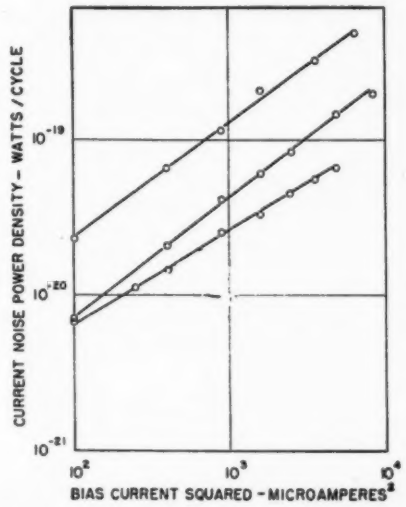


FIG. 3. Variation of the current noise with direct current for three thermistor flakes. The upper two curves were taken at 200 cps and the lower at 20 cps.

resistor was examined. The wirewound unit exhibited no current noise, the noise level being thermal noise only for all values of direct current up to 100 microamperes. The carbon resistor displayed additional noise which was a function of current and had the typical inverse frequency characteristic. It was felt that these studies indicated the system was capable of yielding significant, quantitative information about the current noise of thermistor bolometers.

III. RESULTS

The thermistor flakes studied were five millimeters long, one-half millimeter wide by 20 microns thick and were mounted on glass with apiezon cement. The dc resistance at room temperature was of the order of two megohms. Contact to the two ends was made by various techniques which are discussed below.

The variation of the available current noise power density p defined by

$$p=\frac{v_c^2/B}{4R} \qquad (2)$$

as a function of frequency and bias current was first examined. It was found that the current noise generated by all the samples was very nearly the same, differing only in magnitude. Typical results for several units are shown in Figs. 2. and 3. It is seen that the noise decreases approximately inversely with frequency and directly with direct current. Within the accuracy of the measurements it is found that the current noise may be represented by the expression

$$p=CI^{2b}f^a, \qquad (3)$$

where I is the dc bias current, f is the frequency, and C is a constant. The average of eight samples gives for the

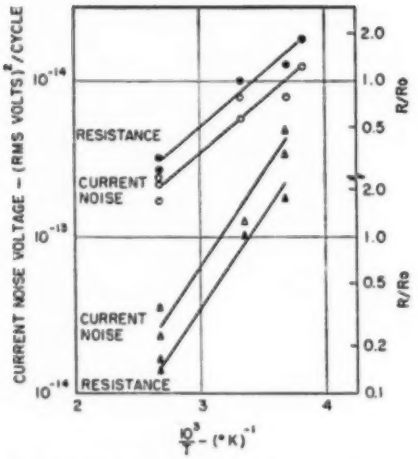


FIG. 4. The change in resistance and current noise voltage as a function of reciprocal temperature for two thermistor bolometer flakes. For clarity, separate resistance scales are used for the two units and R_0 is arbitrarily taken to be the resistance at room temperature.

constants a and b

$$a=-1.10$$

$$b=0.588.$$

The frequency dependence agrees well with that found for other semiconductors,[3] while the variation with current is closer to linear than the more usual square law. In view of the wide variation reported in this effect,[4] this is perhaps not too surprising.

The influence of temperature upon current noise voltage was examined. The noise was measured at constant bias current and for a fixed frequency as the temperature was varied over the limited range that the thermistors will stand without breaking due to differential expansion of the mounting cement. The dc resistance of the flakes was simultaneously measured. The results are shown in Fig. 4 for two flakes. It is seen that the variation of both the resistance and the open-circuit mean-square current-noise voltage is exponential with temperature. Although the accuracy of the results is somewhat poorer than desired, it is indicated that the noise has the same activation energy as the resistance. Thus, the power density calculated from Eq. (2) is found to be independent of temperature.

A number of techniques for attaching electrodes to the flakes were investigated to discover which consistently produced the lowest noise units. A criterion of the noise level was taken to be the magnitude of C in Eq. (3) where, for simplicity, the exponents are assumed to be $a=-1$, and $b=\frac{1}{2}$. Typical values of C for several of the electrodes are given in Table I.

Contacts produced by using platinum bright and platinum mush,[9] which have proven satisfactory when used on the standard thermistor flakes, were found to yield noisy thermistor units in this case. Evaporated platinum electrodes applied under high vacuum are seen to be about the same but evaporated gold contacts are considerably better.

A major reduction in the current noise level was achieved when the evaporated electrodes were applied through a potential difference. In this technique the flake is made 500 volts negative with respect to the evaporating filament during vaporization of the metal. Evaporated platinum applied in this fashion produces units as good as the evaporated gold alone, and the evaporated gold contacts using the potential difference are found to have a noise level smaller by a factor of ten than any other electrodes. The lowest curve of Fig. 2 represents the current noise observed from a unit produced in this fashion.

The evaporated gold potential difference method consistently produces flakes having the low noise levels given above. All attempts to date to improve upon this technique have failed. Raising the potential to 5000 volts does not lower the noise level. Also, an attempt to clean the surface of the flakes by means of a glow discharge before applying the metal was not successful.

TABLE I. Current noise levels of thermistor flakes having different electrodes.

Electrode technique	C (watts/microampere)
Platinum bright	3.3×10^{-17}
Evaporated platinum	3.1×10^{-17}
Evaporated gold	1.3×10^{-20}
Evaporated platinum at 500 v	1.9×10^{-20}
Evaporated gold at 500 v	2.7×10^{-21}

IV. SUMMARY

The additional electrical noise which appears in iron oxide thermistor bolometer flakes when a direct current is passed through them is shown to be similar in frequency response and current dependence to the current noise observed in other semiconductors. The mean-square current-noise voltage varies exponentially with temperature in a manner to make the noise power density independent of temperature. Evaporated gold electrodes applied through a potential difference have produced flakes having the lowest noise levels of those examined.

[9] Supplied by Hanovia Chemical and Mfg. Company, East Newark, New Jersey.

Diffusion of Pulsed Currents in Conductors

L. M. VALLESE
Polytechnic Institute of Brooklyn, Brooklyn, New York
(Received August 6, 1953)

The density distribution and the equivalent depth of penetration of pulsed currents produced by a plane wave incident normally upon a plane conductor are investigated. In particular the cases $i(t)=I_0p(t)$, $[I_0/(t)^{\frac{1}{2}}]p(t)$, $I_0e^{-\alpha t}p(t)$, where $p(t)=U(t)-U(t-T_0)$, are considered.

1. INTRODUCTION

ONE of the problems connected with the pulse technique in radio communications concerns the investigation of the diffusion of transient currents into conductors and in particular the evaluation of the corresponding equivalent depth of penetration. In general such investigation is involved, but simplifies considerably in the case of a plane *TEM* wave, normally incident upon a plane surface which is a boundary between a perfect dielectric and a good conductor, both extending to infinity in all other directions. In considering this case we shall assume that the current associated with the wave is of type $i(t)=I_0f(t)\cdot[U(t)-U(t-T_0)]$ with $f(t)$ bounded, and in particular $f(t)=1$, $f(t)=1/(t)^{\frac{1}{2}}$ (with $t>0$), $f(t)=e^{-\alpha t}$. As special cases also $i(t)=\delta(t)$ and $i(t)=\delta^n(t)$ will be considered.

2. DISCUSSION

From Maxwell's equations, neglecting the displacement current, the diffusion of the wave into a conductor is described by the partial differential equations

$$\nabla^2\bar{H}=\frac{1}{h^2}\frac{\partial\bar{H}}{\partial t},\qquad \nabla^2\bar{J}=\frac{1}{h^2}\frac{\partial\bar{J}}{\partial t},\tag{1}$$

where $\bar{J}$ is the current density, $h=1/(\sigma\mu)^{\frac{1}{2}}$ (σ conductivity, μ permeability of the material; for copper $h^2=0.0142$ m/sec). In accordance with our assumptions, we can refer Eqs. (1) to rectangular coordinate axes, for which the boundary surface has equation $z=0$ and z is taken positive inside the conductor. In particular, if the magnetic field H is parallel to the surface of incidence, for example, $\bar{H}=H_y\bar{j}$, there follows $\bar{J}=J_x(z,t)\cdot\bar{i}$ and Eqs. (1) reduce to the scalar equations

$$\frac{\partial^2H_y}{\partial z^2}=\frac{1}{h^2}\frac{\partial H_y}{\partial t},\qquad \frac{\partial^2J_x}{\partial z^2}=\frac{1}{h^2}\frac{\partial J_x}{\partial t}.\tag{2}$$

The current flows in a direction perpendicular to $\bar{H}$ and is distributed inside the conductor if this has finite conductivity. The total current for a strip of unit width in the direction of the y axis is

$$i(t)=\int_0^\infty J(z,t)\cdot dz=H_y(0,t).$$

By application of the Laplace transform method the second of Eqs. (2) provides

$$J(z,s)=\frac{(s)^{\frac{1}{2}}}{h}I(s)\exp\left[-\frac{z}{h}(s)^{\frac{1}{2}}\right],\tag{3}$$

where $I(s)$ is the Laplace transform of $i(t)$. From mathematical tables it is known that

$$\mathfrak{L}^{-1}\{s^{(n-1)/2}\exp[-a(s)^{\frac{1}{2}}]\}=\frac{1}{2^{n/2}(\pi)^{\frac{1}{2}}t^{(n+1)/2}}\times\exp\left(-\frac{a^2}{4t}\right)\cdot He_n\left[\frac{a}{(2t)^{\frac{1}{2}}}\right],\tag{4}$$

where $He_n(x)$ is the Hermite polynomial of order n.

$$He_n(x)=x^n-\binom{n}{2}x^{n-2}+1\cdot3\binom{n}{4}x^{n-4}-1\cdot3\cdot5\binom{n}{6}x^{n-6}+\cdots.$$

In particular

$$He_0(x)=1,\quad He_1(x)=x,\quad He_2(x)=x^2-1,\ \cdots.$$

In order to utilize the result (4) one can write Eq. (3) as follows:

$$J(z,s)=s^{(n-1)/2}\exp\left[-\frac{z(s)^{\frac{1}{2}}}{h}\right]\cdot\frac{I(s)}{h}s^{(2-n)/2}.\tag{5}$$

Letting

$$\psi(t)=\mathfrak{L}^{-1}\left\{s^{(n-1)/2}\exp\left[\frac{-z(s)^{\frac{1}{2}}}{h}\right]\right\},\qquad \varphi(t)=\mathfrak{L}^{-1}\left[\frac{I(s)}{h}s^{(2-n)/2}\right],\tag{6}$$

there follows in general

$$J(z,t)=\int_0^t\varphi(t-\tau)\psi(\tau)d\tau.\tag{7}$$

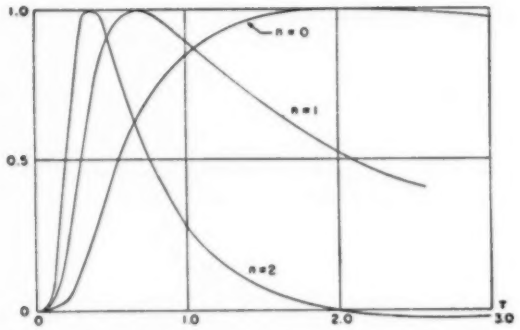


FIG. 1. Plots of the normalized Eqs. (9), (10), and (11).

The difficult evaluation of the integral expression (7) is avoided when

$$\frac{1}{h}I(s)s^{(2-n)/2}=k^2,$$

i.e., when

$$I(s)=k^2hs^{(n/2)-1}, \tag{8}$$

where k^2 is a constant, because in this case $J(z,t)$ is given directly by Eq. (4). Since n in (4) is a positive integer, Eq. (8) corresponds to the following cases of practical interest:

$$n=0,\quad I(s)=k^2h/s,\qquad i(t)=k^2hU(t),$$

$$n=1,\quad I(s)=k^2h/(s)^{\frac{1}{2}},\quad i(t)=k^2h/(\pi t)^{\frac{1}{2}}\cdot U(t),$$

$$n=2,\quad I(s)=k^2h,\qquad i(t)=k^2h\cdot\delta(t).$$

For $n\geq 3$, $i(t)$ is an impulse of half-integral or integral order.[1] Correspondingly the density distribution is

$$n=0,\quad J(z,s)=\frac{k^2}{(s)^{\frac{1}{2}}}\exp\left[-\frac{z(s)^{\frac{1}{2}}}{h}\right],$$

$$J(z,t)=\frac{k^2}{(\pi t)^{\frac{1}{2}}}\exp\left(-\frac{z^2}{4h^2t}\right)\cdot U(t), \tag{9}$$

$$n=1,\quad J(z,s)=k^2\exp\left[-\frac{z(s)^{\frac{1}{2}}}{h}\right],$$

$$J(z,t)=\frac{k^2z}{2(\pi t)^{\frac{1}{2}}t}\exp\left(-\frac{z^2}{4h^2t}\right)\cdot U(t), \tag{10}$$

$$n=2,\quad J(z,s)=k^2h(s)^{\frac{1}{2}}\exp\left[-\frac{z(s)^{\frac{1}{2}}}{h}\right],$$

$$J(z,t)=\frac{z^2-2h^2t}{4h^2(\pi t)^{\frac{1}{2}}t^2}\exp\left(-\frac{z^2}{4h^2t}\right)\cdot U(t). \tag{11}$$

At constant depth z the density reaches a peak at the time t' satisfying the equation $\partial J(z,t)/\partial t=0$, i.e.,

$$\left(-\frac{n+1}{2}+\frac{z^2}{4h^2t}\right)He_n-\frac{nz}{2h(2t)^{\frac{1}{2}}}He_{n-1}=0.$$

There follows

$$\begin{aligned}&\text{for } n=0,\quad t'=z^2/2h^2,\\ &\text{for } n=1,\quad t'=z^2/6h^2,\\ &\text{for } n=2,\quad t'=z^2/6[1\pm(\tfrac{2}{3})^{\frac{1}{2}}]h^2.\end{aligned} \tag{12}$$

These results are illustrated in Fig. 1 where the normalized graphs of Eqs. (9), (10), and (11) are plotted, assuming $u=z/2h$ and $t=\tau u^2$.[2] These are, respectively,

$$\text{for } n=0,\quad \tau'=2.0,\qquad \frac{\epsilon^{-1/\tau}}{0.429(\tau)^{\frac{1}{2}}},$$

$$\text{for } n=1,\quad \tau'=\tfrac{2}{3},\qquad \frac{\epsilon^{-1/\tau}}{0.412(\tau)^{\frac{1}{2}}},$$

$$\text{for } n=2,\quad \tau'=\tfrac{2}{3}[1+(\tfrac{2}{3})^{\frac{1}{2}}]^{-1},\quad \frac{2-\tau}{1.315(\tau)^{\frac{1}{2}}}\epsilon^{-1/\tau}.$$

If we now limit our consideration to the cases $n=0$ and $n=1$ and assume that the current $i(t)$ is a pulse, respectively, of type $i(t)=I_0[U(t)-U(t-T_0)]$ and $i(t)=I_0(t)^{-\frac{1}{2}}\cdot[U(t)-U(t-T_0)]$, the corresponding density distribution is given by Eqs. (9) and (10) for $t\leq T_0$. In particular, for $t=T_0$ a peak value of $J(z,T_0)$ is reached at $z'=h(2T_0)^{\frac{1}{2}}$ for $n=0$ and at $z'=h(6T_0)^{\frac{1}{2}}$ for $n=1$. For $t>T_0$, however, the peak values of $J(z,t)$ beyond the depth z' decrease very rapidly. Therefore, in first approximation, the values of z' may be assumed as corresponding equivalent depths of penetration. In each case the depth is proportional to $(T_0)^{\frac{1}{2}}$; for copper conductor these values are, respectively,

$$z'=0.167(T_0)^{\frac{1}{2}},\quad z'=0.294(T_0)^{\frac{1}{2}}.$$

The corresponding "transient" resistance per unit surface of the conductor is obtained as equivalent dc resistance letting $R_t=1/\sigma z'$. There follows

$$\text{for } n=0,\quad R_t=\left(\frac{\mu}{2\sigma T_0}\right)^{\frac{1}{2}},$$

$$\text{for } n=1,\quad R_t=\left(\frac{\mu}{6\sigma T_0}\right)^{\frac{1}{2}}.$$

In the more general case when Eq. (8) does not apply the computation of the density distribution must be obtained either directly or by application of (7). However, if one is interested only in the values of the

[1] G. Doetsch, *Theorie und Anwendung der Laplace Transformation* (Dover Publications, New York, 1943), p. 298.

[2] The case $n=0$ has been investigated with a somewhat different approach in the following papers: K. W. Miller, Trans. Am. Inst. Elec. Engrs. **66**, 1446–1507 (1947); A. C. Sim, Research report of the British Thomson-Houston Company, Rugby, England (1947).

equivalent depth of penetration, it will be sufficient to find the least root of the equation

$$\frac{\partial J(z,t)}{\partial t}=\varphi(0)\cdot\psi(t)+\int_0^t\frac{\partial\varphi(t-\tau)}{\partial t}\psi(\tau)d\tau=0. \quad (13)$$

Solution of Eq. (13) will provide a relation $t=t(z)$ from which an approximate value of the equivalent depth of penetration z' is obtained. For many applications it will be sufficient to solve Eq. (13) by approximation. If in (6) one chooses $n=2$, then

$$\varphi(t)=\frac{1}{h}i(t),$$
$$\psi(t)=\frac{1}{2t(\pi t)^{\frac{1}{2}}}\left(\frac{z^2}{2h^2t}-1\right)\exp\left(-\frac{z^2}{4h^2t}\right). \quad (14)$$

It has been shown in Fig. 1 that the function $\psi(t)$ of (14) presents a narrow pulse shape with peak at $t'=z^2/\{6[1+(\frac{2}{3})^{\frac{1}{2}}]\}$. This suggests that Eq. (13) might be approximated as follows:

$$\varphi(0)\cdot\psi(t)+\frac{\partial\varphi(t-t')}{\partial t}\cdot\psi(t')=0,$$

i.e., letting $t\geq t'$, $t=\tau u^2$,

$$\frac{1}{u^2\varphi(0)}\frac{\partial\varphi(\tau-\tau')}{\partial\tau}=-\frac{(2-\tau)e^{-1/\tau}}{2.63(\tau)^{\frac{1}{2}}}. \quad (15)$$

For example consider the case $i(t)=I_0e^{-\alpha t}\cdot U(t)$. Equation (15) is written

$$\alpha\exp(-\alpha u^2(\tau-\tau'))=\frac{2\tau}{2.63(\tau)^{\frac{1}{2}}}\exp(-1/\tau). \quad (16)$$

To solve for the least root $\tau=\tau(u)$ with $\tau>\tau'$ one might replace the two members of (16) with their linear approximation, obtaining the system of linear equations

$$\alpha^2u^2\tau+y=\alpha(1+\alpha u^2)\tau',$$
$$\tau+my=\tau_0, \quad (17)$$

where m and τ_0 depend upon the range of interest. For example in Fig. 2 α has been chosen 0.5, $m=1$, $\tau_0=1.3$. From (17) there follows

$$\tau=\frac{\tau_0-\alpha m(1+\alpha^2u^2)\tau'}{1-\alpha^2u^2m};$$

in particular, for $\alpha=0$, $\tau=\tau_0=2$.

The case of an exponentially decayed current $I_0e^{-\alpha t}\cdot U(t)$ can also be analyzed directly, since the current density distribution in complex notation is

$$J(z,s)=\frac{I_0}{h}\frac{(s)^{\frac{1}{2}}}{s+\alpha}\exp\left[-\frac{(s)^{\frac{1}{2}}z}{h}\right].$$

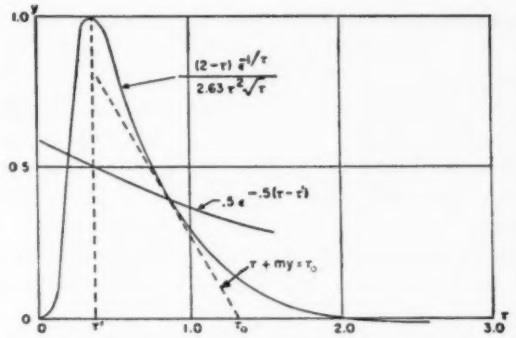


FIG. 2. Graphical representation of Eq. (16) for $u=1$, $\alpha=0.5$.

In the real domain one has

$$J(z,t)=I_0k\left\{\frac{1}{(\pi\alpha t)^{\frac{1}{2}}}\exp\left(-\frac{z^2}{4h^2t}\right)+\mathcal{I}m\exp(-\alpha t+jkz)erfc\left[\frac{kz}{2(\alpha t)^{\frac{1}{2}}}+j(\alpha t)^{\frac{1}{2}}\right]\right\}, \quad (18)$$

where $k=(\alpha)^{\frac{1}{2}}/h$.

The first term of the second member of Eq. (18) represents the distribution of a constant step current, while the second depends upon α and goes to zero when $\alpha\to0$. Unfortunately the complementary error function, which appears in (18) with a complex argument, is not tabulated for all phase angles of the argument. However, an approximate interpretation of the distribution can be obtained by making recourse to the series expansion

$$erfc\xi=1-\frac{2}{(\pi)^{\frac{1}{2}}}\exp(-\xi^2)\left(\xi+\frac{2\xi^3}{1\cdot3}+\frac{4\xi^5}{1\cdot3\cdot5}+\cdots\right), \quad (19)$$

which is uniformly and absolutely convergent everywhere. For small values of ξ one can limit the second member of (19) to the first two terms, i.e.,

$$erfc\left[\frac{kz}{2(\alpha t)^{\frac{1}{2}}}+j(\alpha t)^{\frac{1}{2}}\right]\cong1-\frac{2}{(\pi)^{\frac{1}{2}}}\left[\frac{kz}{2(\alpha t)^{\frac{1}{2}}}+j(\alpha t)^{\frac{1}{2}}\right]\times\exp\left(-\frac{k^2z^2}{4\alpha t}+\alpha t-jkz\right).$$

Correspondingly the second term of (17) can be written approximately

$$I_0k\left[e^{-\alpha t}\sin kz-\frac{2(\alpha t)^{\frac{1}{2}}}{(\pi)^{\frac{1}{2}}}\exp\left(-\frac{k^2z^2}{4\alpha t}\right)\right],$$

and the total distribution is

$$J(z,t)\cong\frac{I_0}{k(\pi t)^{\frac{1}{2}}}(1-2\alpha t)\exp(-z^2/4h^2t)+\frac{I_0(\alpha)^{\frac{1}{2}}}{h}e^{-\alpha t}\sin kz. \quad (20)$$

The examination of (20) shows that for small values of z and t the distribution is sum of one similar to that corresponding to the constant step current, but of smaller magnitude, and of another one consisting of a damped sinusoidal wave which possesses zeros and inversions of sign at depths $z_i=n\pi/k$. For copper conductor the first zero takes place at $z_1=0.182/(\alpha)^{\frac{1}{2}}$ meters.

Let us now compute the equivalent depth of penetration for an exponentially decayed pulse of duration T_0

$$i(t)=I_0\varepsilon^{-\alpha t}[U(t)-U(t-T_0)].$$

Neglecting the last term in (20) one has

$$J(z,t)\cong\frac{I_0}{h(\pi t)^{\frac{1}{2}}}(1-2\alpha t)\exp(-z^2/4h^2t).$$

Equating to zero $\partial J(z,t)/\partial t$ one has, with $u=z/2h$,

$$t-2u^2=-2\alpha t^2-4\alpha u^2 t. \tag{21}$$

Letting $t=T_0$ and solving for u^2 there follows

$$u^2=\frac{T_0}{2}\frac{1+2\alpha T_0}{1-2\alpha T_0},$$

wherefrom the equivalent depth of penetration is

$$z'=h(2T_0)^{\frac{1}{2}}\left(\frac{1+2\alpha T_0}{1-2\alpha T_0}\right)^{\frac{1}{2}}$$

and for small values of α

$$z'=h(2T_0)^{\frac{1}{2}}(1+2\alpha T_0).$$

JOURNAL OF APPLIED PHYSICS VOLUME 25, NUMBER 2 FEBRUARY, 1954

The Effect of Organic Vapor on the Secondary Emission of Phosphors

P. H. DOWLING AND J. R. SEWELL*
Philips Laboratories, Inc., Irvington on Hudson, New York
(Received July 3, 1953)

Experiments are described to show that, when a phosphor surface contaminated with organic vapor is bombarded with electrons, the secondary emission ratio is rapidly reduced. If the bombarding voltage is sufficiently high, this will result in a progressive decrease in potential of the phosphor surface and a corresponding decrease in apparent fluroescent efficiency. The relation between this effect and true "phosphor burn" is discussed as well as its possible connection with the phenomenon of "cross burn" encountered in the cathode-ray tube industry.

IN our early work on cathodoluminescence, we attempted to make measurements using an oil diffusion pump (Octoil *S*) without a cold trap, but with water-cooled baffles, and a demountable cathode-ray tube which could not be baked out. We were disconcerted at the extreme rapidity with which the light output fell off with time as the phosphor samples were bombarded with electrons at a constant voltage. We soon became convinced the trouble was due to organic vapor in the tube and were able to reduce it considerably, even though the tube had joints sealed with Apiezon *W* wax, by using a mercury pump with a liquid nitrogen cold trap. Later we return to the oil pump, this time using a liquid nitrogen trap, but again results were less satisfactory than with the mercury pump. Throughout a series of experiments, the demountable tube presumably accumulated organic vapor during the initial evacuation stages before the cold trap was applied. In a similar way, oil might be accumulated in the demountable tube from the mechanical fore pump. In order to minimize this possibility, most of our experiments have been made with a trap containing solid CO_2 between the fore pump and the diffusion pump. A phosphor which has been in such an atmosphere of organic vapor, and which has been bombarded for a sufficiently long time, will develop on the surface a darkening which is visible in natural light. Considerable decrease in apparent phosphor efficiency occurs, even before the darkening is noticeable. Circumstantial evidence leads us to identify the darkening as carbon formed decomposition of the organic vapor by the electron beam. Other workers have apparently had similar experiences.[1]

Quite recently we became aware that the major effect involved was a reduction, to values less than 1, of the secondary emission ratio δ presumably because of the known low secondary emission of carbon. Hence, the light output measured at a sufficiently high voltage, falls off with time, not primarily because of the absorption of a film of carbon for light or electrons, nor because of an intrinsic loss in efficiency of the phosphor, but because the screen charges negatively, its potential becoming increasingly lower than the second anode voltage. The effect is, therefore, much

* Present address, Sylvania Electric Products Inc., Seneca Falls, New York.

[1] H. W. Leverenz, *Luminescence of Solids* (John Wiley and Sons, Inc., New York, 1950), p. 388.

more noticeable when the phosphor samples are applied to well-insulated glass disks than when they are applied to metal disks connected to the second anode. In the latter case, the conductivity of the sample is usually enough to prevent charging of the screen.

In our experiments, we did not measure δ directly but determined instead the presence of screen charging as evidenced by the departure from linearity of the brightness-voltage curve at higher voltages. For this reason, we usually worked with well-insulated samples. Figure 1 shows results on a phosphor powder settled on a glass disk supported on a glass rod several inches long. An all-glass tube was used which could be baked out except for a small "still" connected to the gun end of the tube, containing about a gram of Apiezon W wax and arranged to be kept cool during the bakeout. After baking out for 16 hours at 400°C and outgassing the gun, the pressure in the tube was $5 \cdot 10^{-7}$ mm. At this point, the data of Fig. 1A were taken, showing little charging of the screen even after nearly 2 hours' bombardment. The wax in the still was now heated to 80°C, while the screen was being bombarded. During this time the pressure ranged from $5 \cdot 10^{-6}$ to 10^{-6} mm. Figure 1B shows that, after 85 minutes of bombardment, there was little effect on screen charging although the efficiency had apparently fallen somewhat, perhaps because of ionic bombardment resulting in a true ion burn. The wax was now held at 50°C for 18 hours with the phosphor unexcited. At this point the pressure was $7 \cdot 10^{-7}$ mm and then, while maintaining the temperature of the wax, the phosphor was bombarded for 90 minutes. Figure 1C shows the relatively enormous effect of this treatment. We interpret the fact that Fig. 1B shows no more evidence of screen charging than Fig. 1A as meaning that the organic vapor was decomposed by the electron beam before reaching the phosphor; i.e., for maximum effect, the organic vapor should actually be adsorbed on the phosphor surface. This conclusion is substantiated by experiments with a demountable tube having wax joints, in which the phosphor could be baked out although most of the tube could not. Here, bombarding a fresh insulated sample immediately after the bake produced little evidence of screen charging. If, however, the same sample were allowed to stand in the evacuated tube for several hours after the bake, then the development of screen charging during bombardment became very evident.

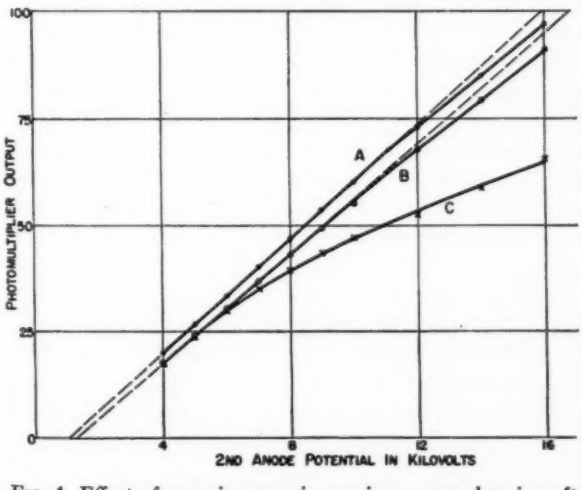


FIG. 1. Effect of organic vapor in causing screen charging after bombardment with a steady beam of electrons 1.0μ amp/cm² at 5.0 kv. All-glass tube mercury pump, $CaMg(SiO_3)_2Ti$ phosphor. A. Initial curve and the coincident curve obtained after 105 minutes bombardment. Wax unheated. B. Wax heated at 80°C while sample was bombarded for 85 minutes. C. Wax heated at 50°C for 18 hours. Then sample bombarded for 90 minutes.

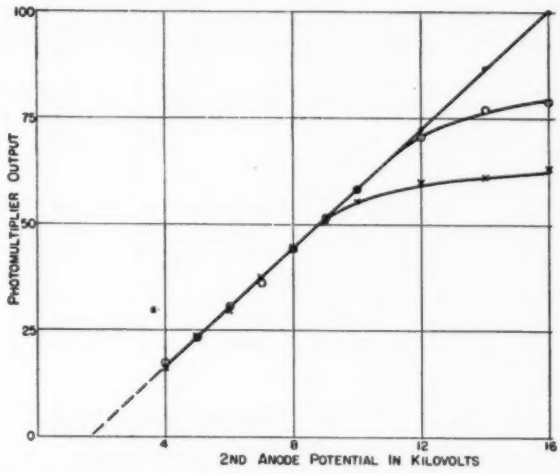


FIG. 2. Effect of phosphor bombardment with a steady electron beam of 1.0μ amp/cm² at 12.0 kv. An all-sulfide P-4 phosphor, sample No. 1. Demountable tube, mercury pump. ● Initial curve. ○ After one hour bombardment. × after two hours' bombardment.

We have some evidence, not entirely conclusive, that some phosphors are more susceptible to the charging effect than others. This may be, for example, because of a difference in the specific adsorption for organic vapor. In a similar way, one might expect the binder used in the screen to have an influence. For the most part, our samples were settled from Kasil solutions in standard fashion, using barium acetate as the accelerator and were thoroughly baked in air at 400°C before installation in the tube. Screens settled from water, however, behaved similarly to those settled from Kasil solutions.

The curves of Figs. 2 and 3 compare the rate of development of charging effects in the demountable tube when the sample is energized by a steady beam to that when it is scanned by a standard television raster of the same average current density. The rate under scanning conditions is much slower, indicating that the decomposition of the organic vapor is not a linear function of the actual current density in the electron beam. Once charging effects have been established by steady beam excitation, scanning excitation does not decrease them, as might occur if the high momentary actual current density in the scanning beam had a tendency to remove carbon from the phosphor surface.

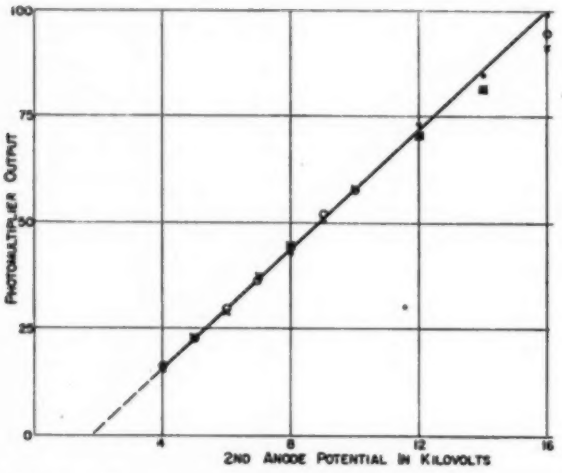


FIG. 3. Effect of phosphor bombardment with a scanning electron beam of 15.0μ amp within 15 cm^2 at 12.0 kv. Same sulfide phosphor as in Fig. 2, sample No. 2. Data taken immediately after that of Fig. 2 in same tube. ● Initial curve. ○ After one hour scanning. × After two hours scanning.

Throughout these experiments, measurement of light output was always made with a steady electron beam.

The phosphor surface contaminated with organic vapor may be subjected to a number of treatments without materially alternating screen charging induced by electron bombardment. These include baking in vacuo at 400°C and exposure to air at room temperature, either dry or at ordinary room humidities. Baking in air at 400°C, however, returns screen charging to its original amount. In this case, if the sample had developed visible darkening during the bombardment, the bake also removes the darkening. These reactions are consistent with the idea that the charging effects discussed above are due to deposition of a thin layer of carbon.

It may be emphasized that, although under some circumstances the two effects may be confused because of a common tendency to lower phosphor efficiency, these charging effects are quite distinct from true phosphor burn caused by the impact of charged particles having considerable energy and resulting in a decrease in the intrinsic efficiency of the phosphor. Aside from the differences obvious from the above paragraphs, the two effects may produce losses in phosphor efficiency which react differently to changes in the voltage of the electron beam used for efficiency measurement. A true phosphor burn which has the nature of a surface dead layer[2] will be most apparent, percentagewise, at low voltage. On the other hand, unless it is very thick, the effect of the carbon layer is most apparent at high voltages.

There is no reason, of course, why the two effects cannot exist simultaneously. For example, a phosphor screen contaminated with organic vapor and bombarded by electrons, whose accelerating voltage is sufficiently high, will reach a condition where the screen potential is below this accelerating potential. When this occurs, the screen will be bombarded by any positive ions in the vicinity and will thus suffer a true ion burn in addition to the charging effect caused by the organic vapor. This mechanism will operate, of course, in the absence of organic vapor whenever the screen potential is below the voltage of the electron beam. It has been suggested by others[3] as the fundamental mechanism responsible for the phenomenon of the "cross burn" which develops in some television tube screens. If this explanation is correct, the presence of organic vapor should accelerate the development of cross burn.

We were readily able to show, in a demountable tube with wax joints, the distinction between charging effects and a true phosphor burn which had resulted in a surface dead layer. When the electrostatically focused electron beam was deflected with a magnet, marked charging effects always developed in an area coincident with the area of the electron beam. The associated loss in phosphor efficiency was very pronounced at high voltage as in Fig. 1C. Bombardment was continued until the darkening of the area was visible in natural light. The sample was now removed from the tube and baked in air at 400°C. This treatment removed all visible trace of the bombardment, but when the sample was replaced in the tube and examined under the electron beam without further treatment, the originally bombarded area was found to have a lower fluorescent efficiency than the surrounding area. Now, however, the lower efficiency was most pronounced at quite low voltages; at high voltages, even up to 20 kv, the originally bombarded area was completely indistinguishable, to the eye, from its surrounding regions.

Although it was slower to develop, an area of low fluorescent efficiency and about equal in extent to that of the electron beam appeared, during the experiment described in the preceding paragraph, nearly on the axis of the electron gun. Qualitatively, it behaved in all respects like the area under the electron beam. Although the details of its origin are not entirely clear, this area must be associated with a beam of nearly undeflected negative ions accelerated to the second anode potential. The fact that it also showed marked charging effects may mean that negative ions are also effective in decomposing adsorbed organic vapor.

[2] P. H. Dowling and J. R. Sewell, J. Electrochem. Soc. **100**, 22–33, (1953).

[3] For example: private communication from H. A. Klasens, Philips, Eindhoven.

Design and Calibration of Stagnation Temperature Probes for Use at High Supersonic Speeds and Elevated Temperatures

EVA M. WINKLER
U. S. Naval Ordnance Laboratory, White Oak, Silver Spring, Maryland
(Received July 13, 1953)

A stagnation temperature probe with a single platinum-coated silica shield was built for use at high supersonic speeds and elevated temperatures. For large Reynolds numbers the temperature recovery factor of the probe reaches 0.998. It decreases with decreasing Reynolds number and increasing temperature. At temperatures up to 700° K the major thermometric losses of the probe are conduction losses. One single calibration curve is obtained for any one probe by relating the calibration data to the flow conditions inside the probe.

INTRODUCTION

FOR temperature surveys in the NOL Hypersonic Wind Tunnel No. 4[1] a stagnation temperature probe has been designed which is specially suited for use at elevated temperatures. Usually, temperature probes for use at supersonic speeds have two or more vented shields around the sensing thermocouple.[2] These shields serve to direct the flow and minimize the heat loss from the air sample the temperature of which is to be measured. For boundary-layer studies in the hypersonic wind tunnel a small probe, of the order of 1-mm outside diameter, is desirable in order to obtain a sufficient number of data through the boundary layer. Such a small probe is easier to assemble, if it can be built with a single shield around the thermocouple.

Since material and finish of the shields, length and diameter of the thermocouple wires, as well as vent area and size of the probe, are parameters affecting the temperature recovery of the probe, it was felt that a single-shielded probe may be made to perform in a manner comparable to multiple-shielded probes by proper selection of these parameters, in particular by not using metal as the shield material as is commonly done for probes used at elevated temperatures.

The probe described in the following is still rather large, 6-mm outside diameter. This size was selected to have a probe large enough for easy handling during the preliminary investigations. The probe is, however, small enough to obtain a number of data points in the nozzle wall boundary layer of the hypersonic wind tunnel.[3]

DESCRIPTION OF THE PROBE

The design of the probe is shown in Fig. 1. The shield is made of silica with both surfaces platinum coated to reduce conduction losses through the shield and radiation losses from the shield surface to cooler surroundings. The shield is cemented to the stainless steel holder with Technical B Copper Cement.* Two vent holes provide a continuous replacing of the air inside the probe. The thermocouple is of 30 gauge (Brown and Sharpe Gauge) iron-constantan wire. It is fiberglas insulated, and sealed into the silica support with Technical B Copper Cement. The exposed surface of the support is also platinum coated. The size of the vent holes and the length of the thermocouple projection have been varied within limited ranges. Best performance was obtained for a vent-area to entrance-area ratio of 1:5.

PERFORMANCE

To evaluate data from boundary-layer surveys, the performance of the probe has been tested over a wide range of flow conditions, i.e., Mach numbers, temperatures, and Reynolds numbers. The calibration data for one specific probe are plotted, Fig. 2, as temperature recovery factor *vs* Reynolds number. The Reynolds numbers refer to free stream conditions and are based on the entrance outside diameter of the probe. The temperature recovery factor is defined as $(T_i-T)/(T_0-T)$, where T_i is the temperature indicated by the probe, T_0 the stagnation temperature, and T the free stream temperature. The stagnation temperature is set equal to the area-weighted average value of the supply air temperature as obtained from a survey in the

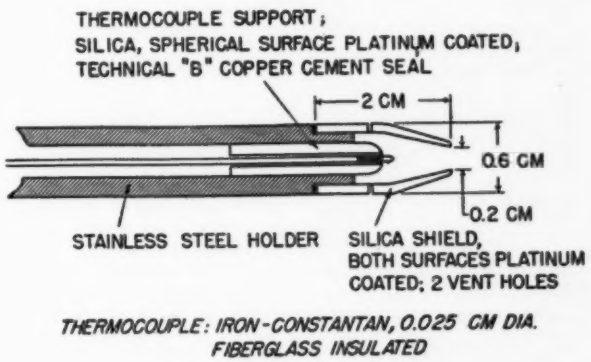


FIG. 1. Stagnation temperature probe.

[1] Peter P. Wegener, J. Aeronaut. Sci. **18**, 665 (1951).
[2] D. L. Goldstein and R. Scherrer, "Design and calibration of a total-temperature probe for use at supersonic speeds," National Advisory Com. Aeronaut. Tech. Note **1885** (1949).
[3] Wegener, Winkler, and Sibulkin, J. Aeronaut. Sci. **20**, 221 (1953).

* Technical B Copper Cement (W. V-B Ames Company, Fremont, Ohio) is a dental cement useful up to 1000° Kelvin.

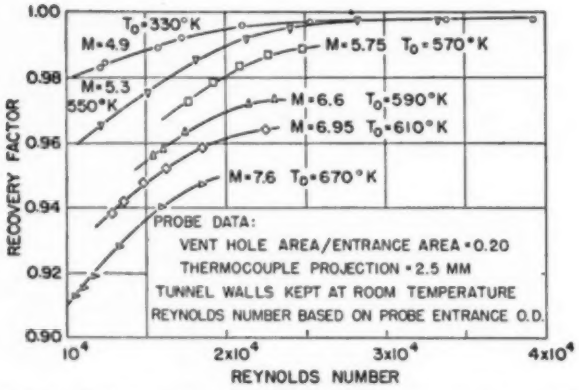


FIG. 2. Variation of probe temperature recovery factor with Reynolds number at various free stream Mach numbers.

nozzle entrance duct.[4] The free stream temperature has been calculated from compressible flow tables.[5] The range of Reynolds number covered for each Mach number is determined by the operational range of the hypersonic wind tunnel. For each Mach number, the supply temperature has been selected high enough to avoid air condensation in the wind-tunnel test section; the supply pressure was then varied in a range so that the set temperature could be maintained with the heater power available. Some data were obtained for Mach numbers below 4.9; they correspond to Reynolds numbers above 40 000, and under these conditions the recovery factor was constant and equal to 0.998. The overall shape of the curves of Fig. 2 seems to be alike for each Mach number; the recovery factor decreases with decreasing Reynolds number because of the relative increase of conduction and radiation errors as compared to the convection heat transfer. For this specific probe and at the temperatures involved, the major losses are conduction losses. For the most unfavorable case, i.e., data in the lower portion of the Mach number 7.6 curve, estimation of the conduction and radiation losses[6] indicate that approximately

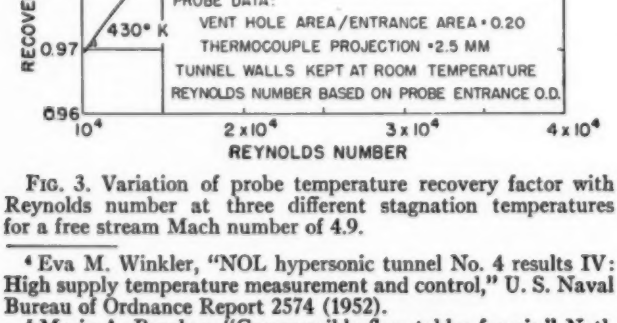


FIG. 3. Variation of probe temperature recovery factor with Reynolds number at three different stagnation temperatures for a free stream Mach number of 4.9.

95 percent of the total losses are due to conduction.

The variation of the recovery factor with temperature alone has been investigated for Mach number 4.9. Three series of tests were made for the same range of Reynolds numbers but at three different temperatures (Fig. 3). With decreasing Reynolds number, the temperature recovery of the probe drops more rapidly as the temperature is increased.

The curves of Figs. 2 and 3 can be combined into a single calibration curve if the measured data are related to the conditions inside the probe instead of to free stream conditions. Plotting the temperature recovery factor as function of the Nusselt number of the flow inside the probe (based on wire diameter) multiplied by the ratio of thermal conductivities of gas and thermocouple wire, the lower curve of Fig. 4 is obtained for the data shown in Figs. 2 and 3. For probes where only the length of the thermocouple projection is changed, this calibration curve shifts as expected when conduction losses are the major losses. The

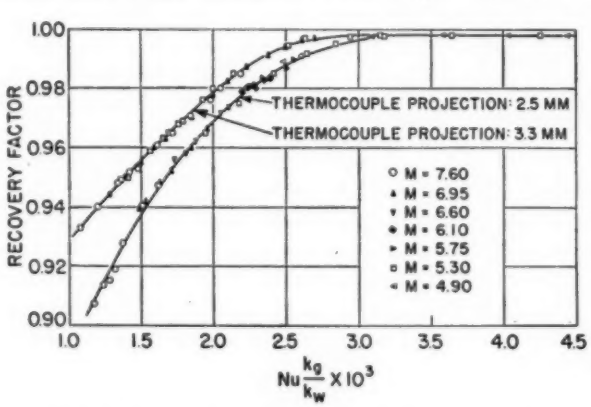


FIG. 4. Summary of calibration data for two probes.

upper curve in Fig. 4 is obtained with an increased thermocouple projection.

Development of this probe is being continued. To reduce conduction losses, the length over diameter ratio of the thermocouple wires will be increased. To reduce radiation losses further, the platinum coating will be replaced by a gold coating and a radiation insulation provided on the exposed thermocouple wires to minimize radiation losses in the probe interior. Finally the dimensions of the probe will be reduced to the desired 1-mm diameter size.

ACKNOWLEDGMENTS

The work described in this paper and its continuation is part of a program sponsored jointly by the U. S. Naval Bureau of Ordnance and the U. S. Air Force.

The author wishes to acknowledge the cooperation of Mr. W. L. Clark and Mr. R. Garren, Naval Ordnance Laboratory, in manufacturing the probes.

[4] Eva M. Winkler, "NOL hypersonic tunnel No. 4 results IV: High supply temperature measurement and control," U. S. Naval Bureau of Ordnance Report 2574 (1952).

[5] Marie A. Burcher, "Compressible flow tables for air," Natl. Advisory Com. Aeronaut. Tech. Note 1592 (1948).

[6] M. D. Scadron and I. Warshawsky, "Experimental determination of time constants and Nusselt numbers for bare-wire thermocouples in high-velocity air streams and analytical approximation of conduction and radiation errors," Natl. Advisory Com. Aeronaut. Tech. Note 2599 (1952).

The Analysis of Quantitative Pole-Figure Data

C. G. DUNN
General Electric Company, Pittsfield, Massachusetts
(Received July 13, 1953)

An analysis of isolated concentrations of poles in a pole-figure plot has been made in terms of a normal bivariate frequency distribution. A definition of pole density is given, and absolute units for measuring pole density are introduced. Simple procedures for determining the standard deviation of the pole-density distribution and for calculating the fraction of crystals in one preferred orientation in a texture consisting of several preferred orientations are given. A method of constructing a quantitative pole figure from a number of single Laue patterns and crystal sizes in coarse-grained material also is given.

INTRODUCTION

THE determination of quantitative pole figures by the Geiger-counter spectrometer method has been described in recent papers.[1–5] The pole figure, which is usually shown in stereographic projection, consists of pole-density contour lines, each a simple multiple of the random orientation value. Pole density may be visualized as a height above an XY plane that lies in the surface of the stereographic-projection reference sphere and the distribution of pole density may be considered as a surface above the XY plane. In statistics this is a bivariate frequency surface.

If the pole figure consists of a sharp single-orientation texture, the bivariate surface for an isolated concentration of poles may take the form of a normal distribution, particularly for the central region of high pole density. This is the type of distribution assumed by DeBarr[6] recently in his consideration of the problem of the required number of crystals to be scanned by the integrating camera method when determining the standard deviation to a given degree of accuracy. It is the purpose of this paper to consider the problem of analysis of quantitative pole-figure data consisting of normal bivariate surfaces and to discuss some recent pole-figure data which agree well with the normal distribution.

THEORY OF NORMALLY DISTRIBUTED POLE DENSITY

Let us consider a single orientation texture with the pole density falling to zero between peak concentration points according to a normal distribution law. Figure 1, showing one quadrant only, may be taken as illustrative of the form of such a pole-density distribution for an isolated concentration of poles. It is a normal bivariate frequency surface with normal distribution curves for any plane perpendicular to the XY plane. Since a curve of constant pole density is an ellipse, we take for simplicity X and Y axes parallel with the major and minor axes of this ellipse. We also take σ_x and σ_y as the standard deviations of the principle normal curves. With these parameters the pole density z is given by

$$z = z_0 \exp\left[-\left(\frac{x^2}{2\sigma_x^2}+\frac{y^2}{2\sigma_y^2}\right)\right], \qquad (1)$$

where z_0 is the peak height and x and y are the coordinates of a direction. The latter are in angular measure.

Integrating over this isolated concentration of poles in the usual way for a normal bivariate frequency surface, we obtain

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} z_0 \exp\left[-\left(\frac{x^2}{2\sigma_x^2}+\frac{y^2}{2\sigma_y^2}\right)\right]dxdy = 1, \qquad (2)$$

when z_0 has the value $1/2\pi\sigma_x\sigma_y$. (Integration limits of $3\sigma_x$ and $3\sigma_y$ instead of ∞ gives essentially the same result.)

Setting the value of the double integral equal to 1 is equivalent to making the element of volume $zdxdy$ under the surface equal to the probability that a crystal of the specimen, selected at random, has a pole of the specified kind in the direction x, y within the solid angle $dxdy$. In terms of weighted numbers (of crystals) the element of volume also gives the fraction of sample having poles within the range $dxdy$ in the direction x, y.

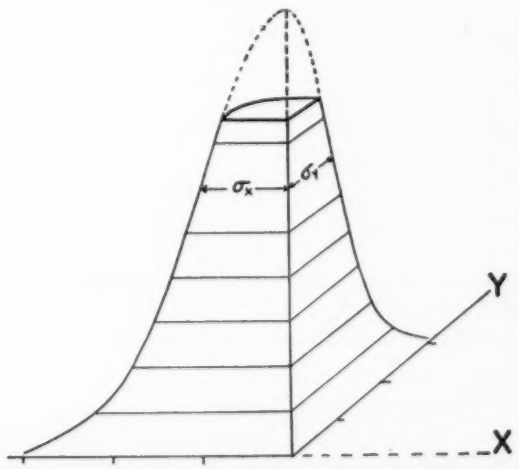


FIG. 1. Quadrant of normal bivariate frequency surface.

[1] Decker, Asp, and Harker, J. Appl. Phys. **19**, 388 (1948).
[2] Am. Soc. Testing Materials Committee E-4 on Metallography. Am. Soc. Testing Materials, Standards, E81-49T (2), 1094 (1949).
[3] J. T. Norton, J. Appl. Phys. **19**, 1176 (1949).
[4] L. G. Schulz, J. Appl. Phys. **20**, 1033 (1949).
[5] A. H. Geisler, *Modern Research Techniques in Physical Metallurgy* (American Society of Metals, Cleveland, 1953), p. 131.
[6] A. E. DeBarr, Sheet Metal Inds. **28**, 603 (1951).

TABLE I. Fraction of sample within two standard deviation units for the entire range of values of the ratio σ_x/σ_y.

Integration limits (Weddle's rule)	σ_x/σ_y	ρ/σ_x	Fraction, $f(2\rho)$
	1	1.414	0.982
0°–30°, 30°–90°	1.5	1.202	0.977
0°–30°, 30°–90°	2	1.118	0.970
0°–18°, 18°–90°	4	1.031	0.960
0°–18°, 18°–90°	10	1.005	0.957
Univariate distribution	∞	1	0.954

Using the distribution function given by Eq. (1) to obtain the average value of x^2+y^2 or the variance, we find the standard deviation ρ is related to the parameters σ_x and σ_y as follows:

$$\rho=(\sigma_x^2+\sigma_y^2)^{\frac{1}{2}}. \tag{3}$$

The standard deviation gives a numerical measure of the sharpness of the pole concentration. For example, over 95 percent of the sample is represented within two standard deviations of the center point. Call this number $f(2\rho)$. To determine the exact value of $f(2\rho)$, which varies with the ratio σ_x/σ_y, we change the limits of integration in Eq. (1) to 0 to 2ρ and 0 to 2π and express the distribution function in polar coordinates r and θ. The radial part can be integrated easily; but the integration over θ is difficult except for the case when σ_x equals σ_y. The θ integration can be done easily, however, by numerical methods (Weddle's Rule, for example). Results are listed in Table I.

ABSOLUTE VALUES OF POLE DENSITY

We may define pole density for any kind of distribution whatever as follows:

$$z=df/d\omega, \tag{4}$$

where df is the fractional element of sample within the element of solid angle $d\omega$.

If x and y in Eqs. (1) and (2) are in radian measure or if $d\omega$ in Eq. (4) is in steradian measure, then z may be said to be in absolute units. Since the unit of pole density has little physical meaning by itself, one may more conveniently think in terms of the product $zd\omega$ or df, which lies in the range zero to one and is either a probability or the fraction of the sample having poles in the element of solid angle.

A stereogram covers a solid angle of 2π steradians when it is a projection of a hemisphere of the reference sphere. In such a range there is more than one pole concentration for a single orientation texture. There are three (100) poles, four (111) poles, six (110) poles, etc., or n (hkl) poles for the general case. Therefore,

$$\int_0^{2\pi} zd\omega=n, \tag{5}$$

which means that the entire sample is counted n times during integration over 2π steradians.

The average pole density, which also is the pole density of a random orientation texture, is $n/2\pi$ in the present units. Average pole density, therefore, depends on the kind of stereogram. It is twice as great for (110) poles as for (100) poles, but this is entirely reasonable.

The number $n/2\pi$, unfortunately, is somewhat unwieldy. A more useful result, applicable to quantative pole figure data that are expressed in terms of the experimentally determined multiplicities of random intensity, is for the pole density of a random orientation to equal n. This result can be accomplished easily by a change in units, which keeps df in Eq. (4) unchanged. With k expressing pole density in the new units we set

$$k=df/d\omega^*, \tag{6}$$

where k now equals $2\pi z$ and $d\omega^*$ equals $d\omega/2\pi$, the new unit of solid angle being the hemisphere.

The pole density of a normal bivariate surface becomes

$$k=k_0\exp\left[-\left(\frac{x^2}{2\sigma_x^2}+\frac{y^2}{2\sigma_y^2}\right)\right], \tag{7}$$

where k_0 equals $1/\sigma_x\sigma_y$ when the entire sample is represented.

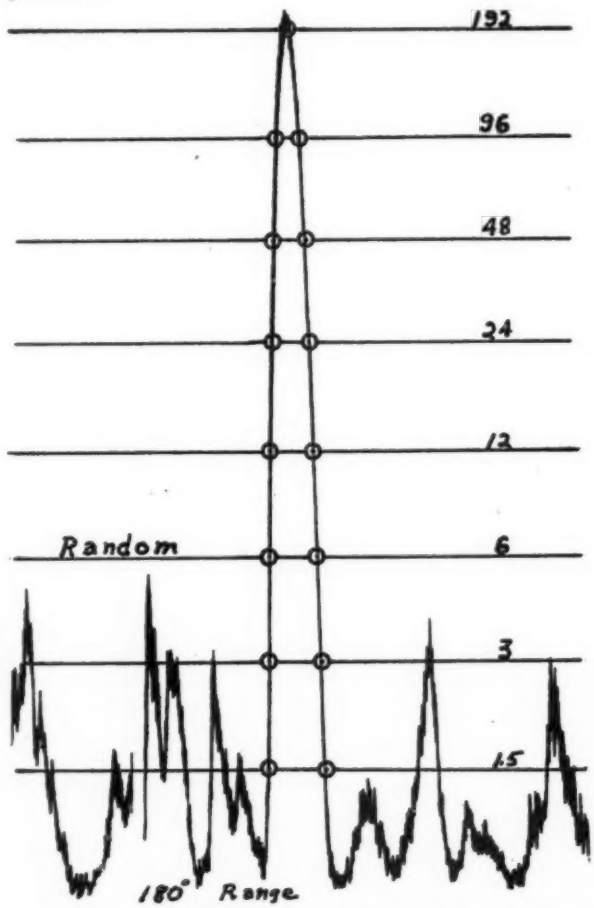


FIG. 2. Geiger-counter spectrometer trace of a fine-grained recrystallized specimen of silicon iron (specimen 2) (see references 7 and 8).

TEXTURE CONSISTING OF SEVERAL PREFERRED ORIENTATIONS

When a texture consists of more than one preferred orientation each with isolated pole concentrations, it is evident that only a part of the aggregate of crystals representing a sample is associated with a single pole concentration. For a normal distribution, we have

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h_A \exp\left[-\left(\frac{x^2}{2\sigma_x^2}+\frac{y^2}{2\sigma_y^2}\right)\right]dxdy/2\pi=f_A, \quad (8)$$

where f_A is the fraction of sample in the preferred orientation A and h_A is the peak height of an isolated pole concentration to which the integration applies.

Equation (8) reduces to the following very useful relationship:

$$h_A\sigma_x\sigma_y=f_A. \quad (9)$$

POLE-FIGURE DATA

In a separate investigation[7] on recrystallization textures and cold-rolled textures produced from single crystals of 3.2 silicon iron, quantitative pole-figure data have been obtained. In this work a transmission Geiger-counter spectrometer method was used to obtain chart recordings of *Mo* K_α diffraction from (110) planes. Figure 2 shows a Speedomax tracing for which, in one narrow range of directions, many crystals were in position to diffract. The heavy horizontal lines were drawn to convert the intensity levels into pole-density levels using the method of direct comparison with a random orientation sample.[5] In terms of relative values, these levels are $\frac{1}{4}$, $\frac{1}{2}$, 1, 2, 4, 8, 16, and 32 times random. On the absolute scale the random orientation pole density is 6.

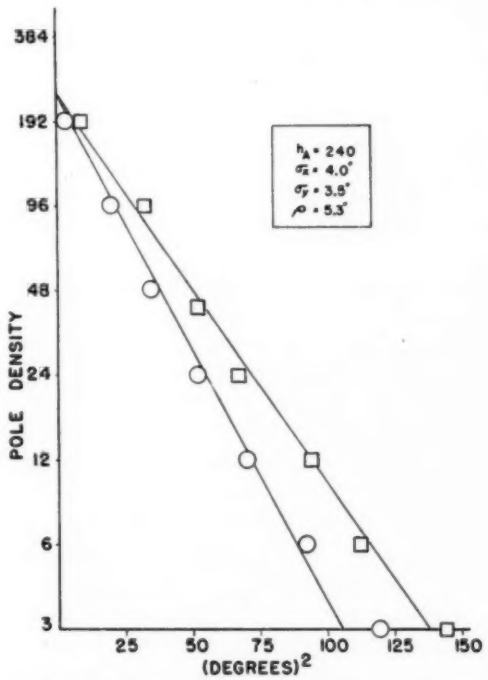


FIG. 3. Variation of pole density with angle from peak position (x or y) for a pole concentration of specimen 1, a cold-rolled single crystal (see references 7 and 8).

[7] C. G. Dunn, Acta Met. (to be published).

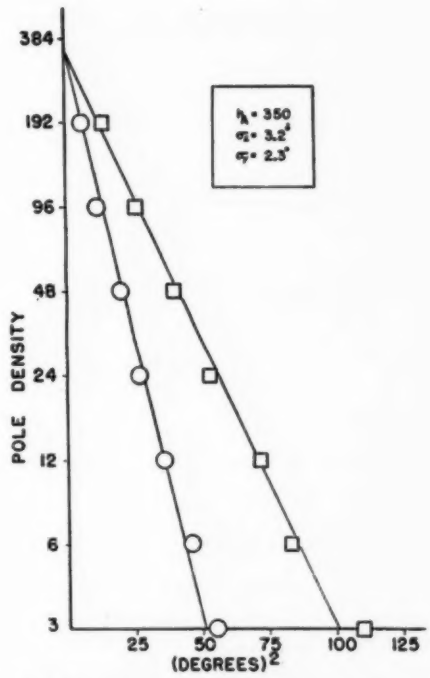


FIG. 4. Variation of pole density with angle from peak position (x or y) for a pole concentration of specimen 1, after recrystallization (see references 7 and 8).

That such a two-dimensional trace follows a normal distribution can be shown by taking h_A equal to 192 and σ_x equal to 2.9° and comparing the calculated value of pole density with the curve. Calculated values are shown by the circled dots. They lie either on or very close to the recorded curve.

To test an actual pole concentration for normal distribution, use is made of the quantitative pole-figure plot. If the constant pole-density contour lines consist of concentric ellipses, which is reasonably true for a number of observed cases, then it is only necessary to read x and y coordinates for each level of pole density and plot results as shown in Figs. 3 and 4 in order to complete the test. The pole density is plotted on a logarithmic scale against the square of the coordinate. Since a relatively good straight line can be drawn meeting at a point that determines k_0 or h_A, one can say that the pole density is normally distributed for these examples. The straight lines also determine σ_x and σ_y; values appear in the figures.

With σ_x and σ_y both in radians the product $h_A\sigma_x\sigma_y$ [see Eq. (9)] for the cold-rolled single crystal of Fig. 3 is 1.02 in good quantitative agreement with a normal distribution for a single-orientation texture. Since the standard deviation is 5.3°, 98 percent of the sample lies within 10.6° of the peak position.

The corresponding product for the recrystallized specimen represented by Fig. 4 is 0.79. It was known

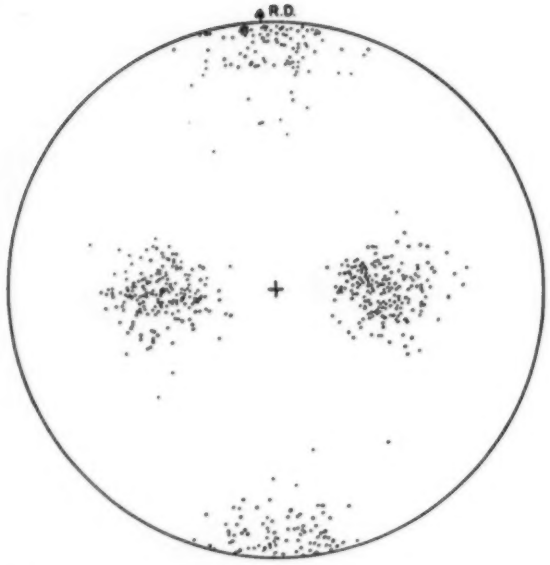


FIG. 5. Stereogram of (100) poles of coarse-grained specimen (from unpublished data).

from the complete pole-figure charts that the strong preferred orientation obtained was accompanied by some minor preferred orientations. These would have to produce an average pole density of 1.2 to account for the missing 20 percent. An estimate, however, would place these contributions at no more than about 5 percent. Also magnetic torque data[8] expressed in terms of relative torque peak with a (110) [001] crystal as 1 showed values over 0.90. It, therefore, seems reasonable to conclude, if the data are sufficiently accurate, that the method of plotting is in error on the low side.

The reason for the low value may be due to the efficiency of x-ray diffraction. The random orientation sample (powdered iron) and the cold-rolled sample may both be more efficient than the recrystallized sample because of greater crystal perfection in the latter case. If this is the source of error, the problem can be solved by cold rolling recrystallized samples a small amount to destroy the extra perfection of the crystals. However, since deformation tends to alter the texture, use of a correction factor instead might be a preferable procedure.

PRODUCTION OF A QUANTITATIVE POLE FIGURE FROM SINGLE-CRYSTAL LAUE PATTERNS

Taken from unpublished work as illustrative for the present procedure, Fig. 5 shows a plot of (100) poles for a disc specimen of silicon iron consisting of 196 crystals and having a relative torque peak of 0.83. Qualitatively, the texture is clearly (110) [001], but in the form shown lacks quantitative characteristics.

The data may be made quantitative by giving each pole a weight according to size of crystal involved, assembling such data to calculate pole density according to Eq. (6) (elements of solid angle of about 0.05 steradians in size are satisfactory), and finally plotting constant density contour lines.

For purpose of analysis, however, one need not make the plot of contour lines. For example, the result shown in Fig. 6 is obtained by taking the pole-density data for the pole concentration in the rolling direction after assuming that constant pole-density contour lines can be taken as circles. It is to be noted that the plotted region can be represented quite well by a straight line. Also application of Eq. (9) gives a value of 0.96, which is in good agreement with a normal distribution for an isolated concentration of poles of a single orientation

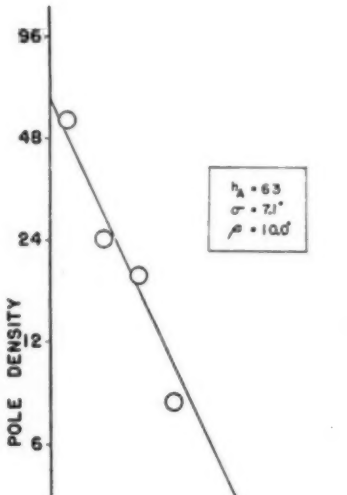


FIG. 6. Variation of pole density with angle of deviation from the rolling direction for a coarse-grained sample of silicon iron (from unpublished data).

texture. The standard deviation calculated from the curve is 10.0° and calculated directly from the original data is 10.9°. The difference is only slightly beyond the 5 percent value of the standard error $[1/(2N)^{\frac{1}{2}}$, where N is 196, if we can consider each crystal with equal weight in calculating the error] and may be the result of some departure from the normal distribution at large angles of deviation from the center.

If a distribution is reasonably close to a normal one, the standard deviation gives a good quantitative measure of the distribution of pole density.

ACKNOWLEDGMENT

The author wishes to thank Dr. A. H. Geisler for discussion of the manuscript, particularly the part dealing with pole-density units.

[8] C. G. Dunn, Acta Met. 1, 163 (1953).

JOURNAL OF APPLIED PHYSICS VOLUME 25, NUMBER 2 FEBRUARY, 1954

The Plastic Deformation of Ammonium Dihydrogen Phosphate

PAUL L. SMITH AND EDWARD I. SALKOVITZ
Metallurgy Division, Naval Research Laboratory, Washington, D. C.
(Received July 18, 1953)

Plastic deformation was produced at about 100°C in single crystals of ADP which is normally considered to be brittle. The following four specimen orientations were studied by loading as a simple beam at temperatures from 90° to 115°C: (a) loaded along [110] axis normal to the (001) face (this orientation being the one of practical importance for all piezoelectric applications); (b) loaded along [100] axis normal to (001) face; (c) loaded along [010] axis normal to (100) face; (d) loaded along [001] axis normal to (100) face. Specimen (a) deformed more readily than the others. X-ray studies were made before and after deformation. Back reflection Laue patterns revealed no difference before and after deformation (i.e., no asterism was detected). A modification of the Berg-Barrett grazing angle technique proved very satisfactory and revealed that structural fragmentation had occurred in these crystals.

INTRODUCTION

AMMONIUM dihydrogen phosphate (usually known as ADP) is a widely used piezoelectric crystal which at room temperature behaves as a very brittle material. It was found,[1] however, that plastic deformation can occur in this material below 100°C which is well within the temperature limit ordinarily regarded as safe for handling. This deformation does not appreciably alter the piezoelectric or elastic properties.

The purpose of this preliminary paper is to describe some of the qualitative results obtained concerning the deformation and to present some suggestions as to the mechanism of deformation derived from x-ray studies.

Chemically, ADP ($NH_4H_2PO_4$) is water soluble, but contains no water of crystallization, and is quite stable to about 100°C. Near 120°C it loses ammonia rather rapidly. The crystal is of tetragonal symmetry of Class V_d. The symmetry elements are an axis of fourfold alternating symmetry (the z axis), two equivalent two-fold axes (x and y), and two planes of symmetry which intersect in the z axis and bisect the angle between x and y.

For this preliminary work, specimens of four orientations were chosen as follows:[2] (a) zxl 45 (this is the principal cut used for piezoelectric purposes and is commonly referred to as the 45 Z cut), (b) zx, (c) xz, and (d) xy. The last three are all unrotated cuts with the edges of the plate lying along crystal axes.

DEFORMATION TECHNIQUES AND RESULTS

Most of the specimens of the zxl 45 orientation were chosen from a group of commercial crystals whose history was unknown. The other orientations were cut from large blocks of ADP grown at NRL. These were wet-lapped to proper dimensions (1.5 in.×0.5 in. ×0.125 in.) thickness tolerance ±0.001 in. and then given a fine grind and light etch. The etching treatment consisted of 5 minutes at room temperature in a solution of ethylene glycol 80 percent, water 20 percent. This treatment was found to remove all traces of surface working.

Of several methods for deforming the specimens which were tried the most satisfactory for these tests was to mount the specimen as a simply supported beam loaded centrally with 3000 grams. The load was supported by flexible wires hung on a $\frac{1}{16}$-in. rod which was in contact with the specimen. The crystal was brought to temperature, loaded, and held for two hours. The angle at the apex of the bent specimen was taken as a measure of the deformation.

Under this treatment measurable deformations of the zxl 45 cut were found to occur as low as 90°C (angle of bend less than 1°) and large deformations (angle of bend greater than 13°) at 115°C. Figure 1 shows a typical example of a bent specimen. A wide variation in the behavior of individual specimens prevented the establishment of a quantitative deformation-temperature relation.

The deformation takes place almost entirely in the region of application of the load. This was found to be true not only on the centrally loaded specimens but on others loaded at two points $\frac{1}{4}$ of the length from each end. Deformation of uniform curvature could be produced by loading the specimen as a column in compression.

Of the four orientations used in tests at 110°C the

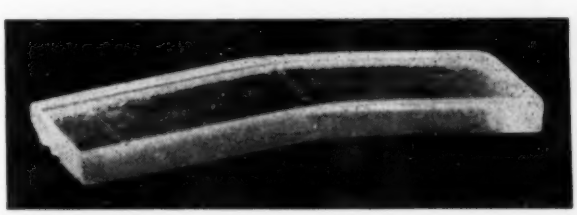


FIG. 1. Bent crystal specimen.

[1] The possibility of plastic deformation in ADP was brought to our attention by Mr. P. N. Arnold of the Sound Division who found curved plates in a transducer which had been opened for repair.

[2] The notation used here for the various cuts is that adopted by the Institute of Radio Engineers. In this notation the first two letters designated, respectively, the crystal axes (x, y, or z) along which the thickness and length are regarded as lying in the unrotated cut. The third letter denotes the edge of the plate (length l, width w or thickness t about which a rotation is made and the number is the number of degrees rotated.

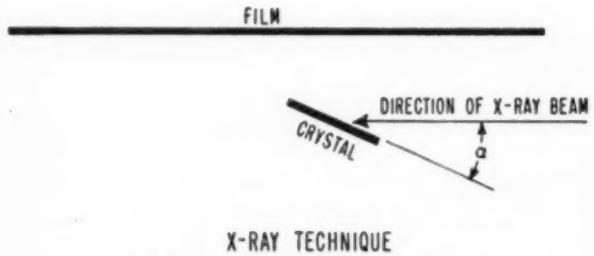


FIG. 2. X-ray grazing angle technique.

zxl 45 was found to deform most readily, the *xy* the least. The *zx* and *xz* were intermediate. No slip lines were observed in any of the tests.

To check the effect of the deformation on the electromechanical properties, the resonant and antiresonant frequencies of a specimen were measured before deformation, after deformation, and after a reversed deformation had brought the specimen back to approximately its original shape. No significant change was found indicating that both the elastic and piezoelectric properties were unchanged by the deformation.

X-RAY STUDIES

Back reflection patterns showed no significant difference between the deformed and undeformed crystals. A modification of the Berg-Barrett grazing angle technique, however, proved very satisfactory in demonstrating the nature of the deformation. In this technique the crystal was mounted on a simple goniometer (see Fig. 2) in front of the x-ray tube. A fluorescent screen was placed above the crystal and essentially parallel to the x-ray beam. The crystal was rotated in the darkened room until x-ray reflections were seen on the screen. The goniometer was then locked in position and the screen replaced by a fine grain x-ray film. Chromium unfiltered radiation and type M film were used in these experiments. With a 0.33-mm diameter pinhole, the exposure times were 50 milliampere-hours at 35 kv when the specimen was about an inch from the window and the film about an inch above the specimen.

RESULTS

X-Ray Patterns

Figure 3 shows a pattern obtained from an undeformed *zxl* 45° crystal in the manner described. The x-ray beam was in the (110) plane of the crystal and made an angle of about 15° with the [110] or *P* axis lying in the face of the crystal. The spots are seen to be relatively sharp and with little structure.

Figure 4 shows a pattern obtained from an *xz* cut which had been bent about the *y* axis. The x-ray beam impinged on the crystal perpendicular to the bending axis. The resulting pattern shows spots containing striae of maxima and minima of intensity, these striae being parallel to the bending axis.

In Fig. 5 is a pattern from a *zxl* 45° (45° *Z* cut) crystal which had been bent about the [110]. With the geometrical arrangement as above, the x-ray beam was again perpendicular to this axis. The reflections are

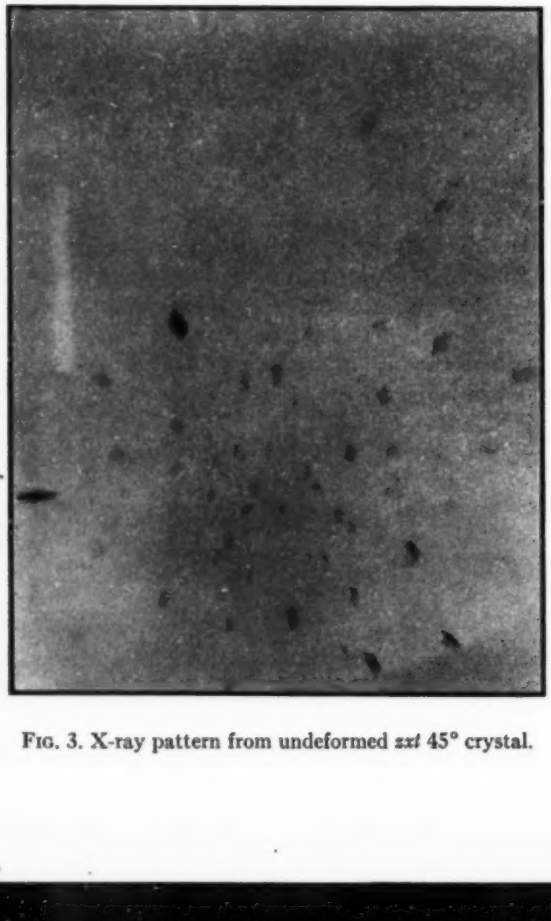


FIG. 3. X-ray pattern from undeformed *zxl* 45° crystal.

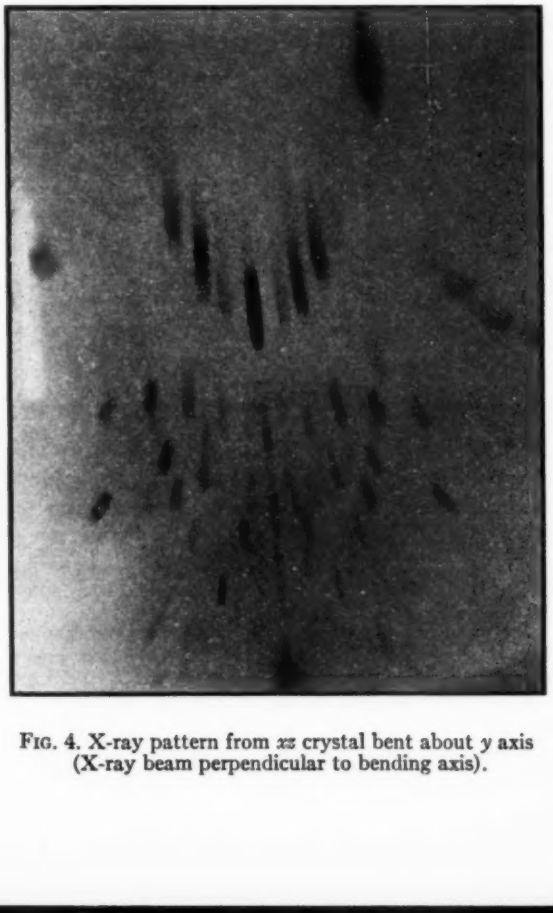


FIG. 4. X-ray pattern from *xz* crystal bent about *y* axis (X-ray beam perpendicular to bending axis).

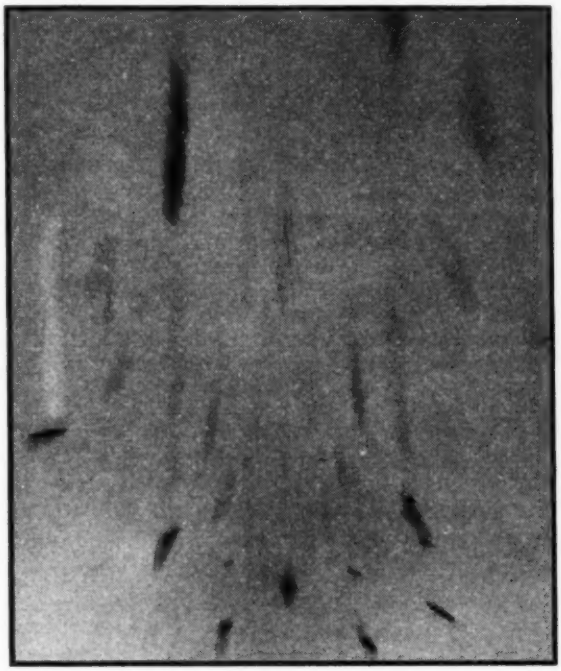


FIG. 5. X-ray pattern from *zxt* 45° crystal bent about [110] (x-ray beam perpendicular to bending axis).

FIG. 6. X-ray pattern from same *zxt* 45° crystal (x-ray beam parallel to bending axis).

elongated and again crossed by maxima and minima roughly parallel to the bending axis.

Figure 6 shows another pattern from the same *zxt* 45° crystal. Here, however, the crystal had been rotated in the azimuthal plane so that the x-ray beam was *parallel* to the bending axis instead of perpendicular to it. It can be seen that very little structure appeared in the spots. Compare with Figs. 3 and 5.

As a result of these patterns and others we are led to believe that in the *zxt* 45 crystals, blocks of material rotated uniformly about the bending axis giving rise to the maxima and minima of Fig. 4. Further, these blocks must have been filament-like, extending across the width of the crystal parallel to the bending axis. Measurements of the spacings of the maxima indicate that the blocks were misoriented relative to one another by 1.3°. This value incidentally is of the same order of magnitude as found for the misorientation of mosaic blocks in deformed metals.

Crystals with other orientations were studied and yielded the following results:

(a) When a *zy* crystal was loaded along the *x* or [100] axis, the x-ray pattern consisted of well-defined maxima and minima parallel to the bending axis. Again we interpret this to mean that in the deformed crystal, blocks of material were tilted about the [100].

(b) When a *yz* crystal was loaded along *x* the resulting x-ray pattern was similar to that obtained in (a), as expected.

(c) When an *xy* crystal was loaded along *z*, [001] the resulting x-ray patterns consisted of diffuse streaks with only one or two maxima appearing parallel to the bending axis. As yet we do not have a clear picture of the manner in which deformation occurred in this case. It should be noted that for a given load the *xy* crystal bent to a far lesser extent than the other crystals discussed.

The crystal mentioned previously, which had been bent and then straightened, gave a pattern which showed that the mosaic structure had been "frozen in." It will be recalled that the piezoelectric constants had not changed.

Other orientations are being studied. It is planned to obtain stress-strain curves in both tension and shear as a function of temperature and to study the effect of plastic deformation on physical properties such as resistivity which should be sensitive to changes in microstructure.

ACKNOWLEDGMENT

Acknowledgment is due Mr. Walter Sadler for taking the x-ray photographs and to Mrs. R. L. Acton for the careful preparation of the oriented specimens.

X-Ray Noise Observation Using a Photoconductive Pickup Tube

A. D. Cope and A. Rose
Radio Corporation of America, RCA Laboratories Division, Princeton, New Jersey
(Received July 18, 1953)

A television pickup tube arrangement using a Vidicon for the direct recording of x-ray images is described. The transmitted x-ray picture shows noise characteristics that are clearly those of the absorbed x-ray photons. The internal conversion gain (number of carriers generated per absorbed x-ray photon) is estimated by two independent methods, one of which is based on a measurement of x-ray noise. Both methods give a gain of about 500 for a 100 kv x-ray photon. The agreement between the two methods is evidence for the validity of the simple electron-multiplier model for the noise current in the photoconductive selenium target. The appearance of x-ray noise means that all the information contained in the absorbed fraction of the x-ray beam is being transmitted. Thirty pictures per second are transmitted when irradiating the Vidicon with a 5 ma, 100 kv x-ray beam at a distance of two feet. Wires of one mil diameter were clearly observable.

THE photoconductivity of evaporated layers of amorphous selenium and of antimony trisulfide and their use in television pickup tubes of the Vidicon type have already been reported.[1–3] Tests made early in the investigation of these targets showed that they would record a picture as the result of direct exposure to x-radiation. The sensitivity of the direct exposure did not equal that obtainable by optical pickup from a fluorescent screen because of the small fraction of the x-rays absorbed by the thin Vidicon targets. The fluorescent screen absorbs and converts practically all the x-radiation to visible light. Selenium targets of 1 to 2 mil thickness deposited on glass face plates 0.020 in. thick showed only about 15 percent absorption of the incident 75 kv x-radiation.

There is, however, both technical and scientific value in knowing how efficiently the photoconductor uses those high energy photons which it does absorb. If each photon were to create several hundred electron-hole pairs, there should be a similar increase in the noise level of the photoconductive signal which would then be observable above the amplifier noise of the camera device.[4] To test this theory, an experimental arrangement was prepared in which a lead shield containing two windows was placed before the face of a Vidicon having a 1 mil thick selenium target. One window was illuminated by x-rays and the second by visible light. The intensity of the light illumination was adjusted until the images reproduced on the kinescope were of equal intensity. Noise (x-ray photon noise) was clearly observable in the x-ray image and was absent in the light image, as is shown in Fig. 1.

The gain factor in the photoconductor, defined as the number of electron charges arriving at the signal plate per absorbed photon, was evaluated by matching the noise in the x-ray generated signal with the shot noise from a temperature limited diode. If each absorbed x-ray photon gave rise to one electron charge in the signal plate of the Vidicon, the gain would be unity. Under these conditions, the mean squared noise current per unit band width in the Vidicon signal plate lead would be

$$\langle I_n^2\rangle_{Av}=2eI, \tag{1}$$

where e is the electronic charge in coulombs and I is Vidicon signal plate current in amperes. In this case I is numerically equal to the number of x-ray photons absorbed in the target per second multiplied by the electronic charge. Equation (1) is simply the shot noise in a temperature limited current of I amp. It is used here on the assumption that the x-ray photons are absorbed randomly in time just as electrons in a temperature limited diode are emitted randomly in time.

If each absorbed x-ray photon gives rise to G electron charges in the signal plate of the Vidicon, the effect is to multiply both the signal current and the rms noise current by the factor G. This is also the well-known result for an electron multiplier; the signal current and rms noise current leaving the cathode of the electron multiplier are both multiplied by the gain of the multiplier.* The mean squared noise current per unit band width in the signal plate lead of the Vidicon is then given by

$$\begin{aligned}(\langle I_n^2\rangle_{Av})_G=G^2(\langle I_n^2\rangle_{Av})&=2eG^2I\\&=2eG(GI)\\&=2eGI_G.\end{aligned} \tag{2}$$

In Eq. (2), $\langle I_n^2\rangle_{Av}$ and I have the same meaning as in Eq. (1), namely, the noise and signal currents appropriate to a gain G of unity. $(\langle I_n^2\rangle_{Av})_G$ and I_G are the corresponding noise and signal currents appropriate to a gain of G between the absorbed x-ray photon current and the observed Vidicon signal plate current.

It is now clear from Eqs. (1) and (2) that if the mean squared noise current per unit band width $(\langle I_n^2\rangle_{Av})_G$ is

[1] Weimer, Forgue, and Goodrich, Electronics 23, 70 (1950).
[2] P. K. Weimer and A. D. Cope, RCA Review XII, No. 3, 314 (1951).
[3] Forgue, Goodrich, and Cope, RCA Review XII, No. 3, 335 (1951).
[4] A. Rose, Advances in Electronics I, 134, 160–163 (1948).

* The fluctuations in the multiplication process itself are neglected here. They are likely to contribute only a small correction to the foregoing relation.

to be matched by the noise from a temperature limited diode, the temperature limited diode current must be (GI_G). That is, the temperature limited diode current must exceed the current in the Vidicon signal plate lead by the gain factor G.

The match between the x-ray noise current and the temperature limited diode noise current was made by observing their envelopes on an oscilloscope, connected near the output stage of the television amplifier. The band width of the amplifier was restricted to one megacycle for convenience of measurement.† First, an observation was made of the envelope of the x-ray noise current on the oscilloscope, and the value of the x-ray signal current in the signal plate lead of the Vidicon was recorded. Second, a noise diode was substituted for the Vidicon and its current adjusted so that the envelope of its noise, as observed on the oscilloscope, matched that already observed for the x-rays. The noise diode current was noted. The gain G was then computed by taking the ratio of the noise diode current to the x-ray signal current.

FIG. 1. Comparative noise caused by illumination by x-rays and by visible light. The upper area was illuminated by visible light to give a signal of equal intensity to that of the lower area which was illuminated by x-rays. The coarse mottling in the x-ray picture which is absent in the visible light picture is the result of x-ray photon noise.

† It is interesting to note that the x-ray noise is stored for the television frame time of 1/30 sec. This means that the x-ray noise in the stored charge pattern corresponds to low frequency noise of 30 cy/sec band width. By the scanning process this charge pattern is picked up at a high rate of speed and observed at an expanded band width of several Mc/sec.

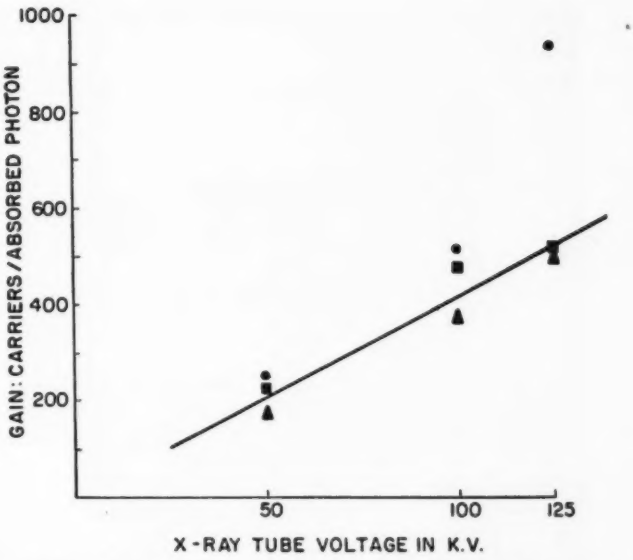


FIG. 2. Gain, defined as the number of charges arriving at the signal plate per absorbed x-ray photon, as a function of the x-ray tube voltage when irradiating a Se Vidicon. The above gains were computed from the magnitude of noise current in the transmitted x-ray pictures.

The x-ray source was an industrial radiographic unit with an inherent filtration of 0.6 mm of aluminum. In addition, there was an absorbing path of 25 cm of air and of 2.5 mm of Kovar sealing glass. The Vidicon tube with a 1 mil thick selenium target was operated in a standard RCA Industrial Television unit with the lens mounting removed.

The observations were made at a number of different values of x-ray voltages and currents. The rms x-ray noise current and the x-ray signal current both doubled when the x-ray voltage was doubled. An approximate check was also obtained of the proportionality of the rms x-ray noise to the square root of the x-ray generating tube current at a given voltage across the x-ray tube. The gain as a function of x-ray tube voltage is shown in Fig. 2. For a 100 kv photon an average value for the gain is 400.

The gain as recorded in Fig. 2 is the product of the number of carriers generated by an x-ray photon and the number of electron charges arriving at the signal plate per generated carrier. The latter factor is called the quantum yield of the photoconductor. Thus, the number of carriers generated by an x-ray photon is equal to the observed gain divided by the quantum yield of the Vidicon target. At the same operating potential at which the gain was observed to be 400, the quantum yield of the photoconductor for blue light was measured to be 0.8. If one assumes that this quantum yield also holds for the carriers generated

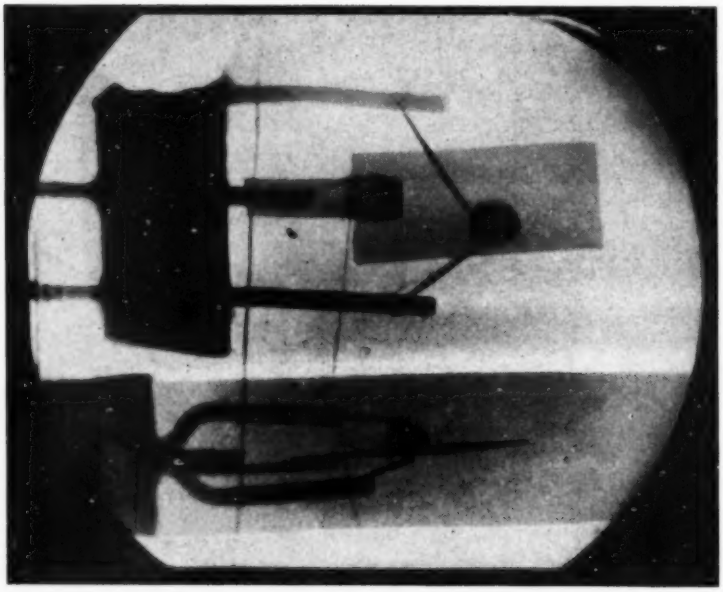


FIG. 3. Photograph of the kinescope showing the picture transmitted when two potted junction type transistors were laid on the face of a Vidicon and irradiated by a 100 kv, 5 ma source of x-rays from a distance of two feet. The two vertical lines are the shadows of fine tungsten wires, a two mil diameter wire on the left and a one mil diameter wire on the right.

by x-radiation, then, the number of carriers generated per absorbed 100 kv photon is 500. Thus, a maximum energy for the formation of an electron-hole pair is 200 electron volts. This value is likely to be lower when proper correction is made for the spread of x-ray energies and for the less than linear relationship of the Vidicon signal current to the x-ray intensity.

An independent check of the number of carriers generated per absorbed x-ray photon was made by comparing the observed photoconductive current through the Vidicon target with the "current" of x-ray photons absorbed in the target. To obtain the latter, the intensity of x-radiation was measured at the Vidicon by means of a Victoreen condenser r-meter. The fraction of x-radiation transmitted by the face plate of the Vidicon was measured to be 30 percent. The fraction of x-radiation absorbed by the one mil thick Vidicon target was taken to be 15 percent from measurements on other layers of selenium. A value of 590 carriers generated per 100 kv x-ray photon was computed in this way. This is to be compared with the value of 500 computed from noise measurements. The agreement between this direct measurement of gain and the "noise" measurement of gain supports the validity of the electron-multiplier model for the x-ray noise currents.

More recent measurements of noise currents in photoconducting cadmium sulfide crystals have lent further support to this simple model. A thin crystal plate of insulating CdS operated as a photoconductive target showed x-ray photon noise when irradiated from a distance of two feet by a 1 ma, 10 kv x-ray beam. A conservative estimate of the number of carriers generated per absorbed x-ray photon was 1000. This would correspond with not more than 10 electron volts for the generation of an electron-hole pair. Also, detailed noise measurements on high gain CdS crystals irradiated with visible light and used as simple photocells were reported to be consistent with the concept of this material as a "noiseless" multiplier.[5]

The ready observation of noise in the x-ray picture means that all the information contained in the fraction of the incident x-radiation, absorbed by the photoconductor, is being transmitted. That is, greater signal-to-noise ratios for a given incident x-radiation can come only through absorbing a greater part of the incident energy. Thicker targets or photoconductors of higher atomic number would achieve this.

There is, however, an advantage in employing thin photoconductive layers of the mil thickness range. Higher resolution is obtained at the expense of sensitivity by the reduction of scattering. It was found, for example, that one mil wires were easily observed, especially since the final picture on the monitor offered a magnification of up to 50 times the original (Fig. 3).

Good pictures at the rate of thirty per second (television scanning rate) were obtained at a distance of two feet from a 5 ma, 100 kv constant voltage source of x-rays. Photographic exposures under the same conditions required several seconds. In principle, x-ray film should also convey all of the information contained in the part of the x-radiation absorbed by the film since, according to Bromley and Herz[6] and Hoerlin,[7] absorbed x-ray photons of greater than 20 kv energy cause at least one photographic grain to be developable. Under these conditions the x-ray noise or graininess should be equal to or greater than the graininess of the film itself and should limit the information recorded.

The authors are indebted to C. J. Busanovich for preparing photographs of the noise phenomena.

[5] Shulman, Smith, and Rose, Phys. Rev. **92**, 857(A) (1953).
[6] D. Bromley and R. H. Herz, Proc. Phys. Soc. (London) **63** Part 2, No. 362B, 90 (1950).
[7] H. Hoerlin, J. Opt. Soc. **39**, 891 (1949).

JOURNAL OF APPLIED PHYSICS VOLUME 25, NUMBER 2 FEBRUARY, 1954

On the Focusing of High-Current Electron Beams

GEORGE R. BREWER
Electron Tube Laboratory, Hughes Research and Development Laboratories, Culver City, California
(Received July 22, 1953)

The focusing of high-current electron beams by means of a system of periodically spaced magnetic lenses has been studied. The trajectories of the electrons under the influence of magnetic and space charge forces have been obtained by the use of an analog computer. Graphs have been prepared which permit the design of a lens focusing system for given values of current, voltage, and beam diameter. The solutions are obtained for the Glaser magnetic lens with longitudinal field variation of the form $B_z=B_0(1+(z/a)^2)^{-1}$. The conditions of operation appropriate to the minimum chromatic aberration are obtained, and the entrance conditions to the lens system discussed.

THE focusing of electron beams by electrostatic or magnetic lenses has been the subject of a large number of papers,[1,2] which, however, have considered electron beams of sufficiently small current density that space-charge effects could be neglected. The effects of space charge in a focused electron beam have also been the subject of several studies.[3,4] However, in these studies the electrons are assumed to enter a field-free drift region with a given initial convergent slope, the details of the lens necessary to produce this initial focusing being omitted.

The purpose of this study is to bring the two aforementioned cases together in a unified treatment of the effects of both the magnetic lens and the space charge on the focusing of electron beams of such current density that space-charge expansion of the beam cannot be neglected.[5]

Magnetic lenses for the control of high-current electron beams have a number of applications in the electron tube field where it is desirable to use long, thin, relatively uniform diameter beams. With the use of such lenses for focusing of the beam, a considerable saving in magnet weight, space, and power consumption over the conventional solenoid can be effected.

To show that one can easily control beams carrying useful current, consider for example, the electron beam in a low-power traveling wave tube. A typical value for the beam perveance in such a tube would be approximately $I/V^{\frac{3}{2}}=0.15\times10^{-6}$. The limiting perveance of a properly focused electron beam injected into a drift tube or helix of diameter D and length l is[6] $I/V^{\frac{3}{2}}=38.6\times10^{-6}(D/l)^2$. If the beam is considered to be focused by n magnetic lenses spaced a distance L apart, $L=l/n$ then the limiting perveance is

$$I/V^{\frac{3}{2}}=38.6\times10^{-6}(Dn/l)^2,$$

so that the typical value of perveance quoted above can be obtained with $n=2$ and $l/D=32$.[7]

The following physical situations will be considered in this analysis:

(a) Focusing of a solid electron beam by a magnetic lens when the focal length is long compared with lens thickness and space charge expansion within the lens is negligible; (b) focusing of a long electron beam by an iterated system of magnetic lenses, taking into account space-charge expansion in the lens, but neglecting overlap of the fields of adjacent lenses; (c) effect of the nearest neighboring lens in case (b); and (d) a simple graphical procedure for finding the envelope of a beam focused by an electrostatic lens, taking into account space charge expansion. The results of the studies of the magnetic lens cases are presented in the form of graphs that permit the design of a periodic lens system for given conditions.

The consideration of the magnetic lens is restricted to the case of the Glaser lens; that is, to axial variation of magnetic field of the form[8]

$$B_z/B_0=(1+(z/a)^2)^{-1}.$$

In all of the equations mks units will be used.

[1] *Advances in Electronics*, edited by L. Marton (Academic Press, Inc., New York, 1950), Vol. II. Zworykin, Morton, Ramberg, Hillier, and Vance, *Electron Optics and the Electron Microscope* (John Wiley and Sons, Inc., New York, 1945).
[2] W. Glaser, Z. Physik **117**, 285–321 (1941).
[3] J. R. Pierce, *Theory and Design of Electron Beams* (D. Van Nostrand Company, Inc., New York, 1949), Chap. IX.
[4] E. E. Watson, Phil. Mag. **3**, 849 (1927); B. J. Thompson and L. B. Headrick, Proc. Inst. Radio Engrs. **28**, 318–324 (1940); D. L. Hollway, Australian J. Sci. Research **A5**, 430–436 (1952); H. Moss, Wireless Engr. 316–321 (July, 1945).
[5] While this paper was in preparation, the work on periodic magnetic focusing of electron beams done at the Bell Telephone Laboratories was described by J. T. Mendel, W. H. Yocum, A. M. Clogston, and H. Heffner at the Conference on Electron Tube Research held at Stanford University June 18–20, 1953. These workers pointed out the practical advantages of the opposing field magnets and presented the results of analysis of focusing conditions in a beam for somewhat different magnetic field conditions than is considered in this paper. Experimental confirmation of their analytical results was also given.
[6] Reference 3, p. 151.
[7] To make this example somewhat more general, a quantity proportional to gain of a traveling wave tube can be written as

$$CN=(KI_0/4V_0)^{\frac{1}{2}}fl/u_0,$$

where C is the Pierce gain parameter, N the number of electronic wavelengths in the tube, $u_0=(2\eta V_0)^{\frac{1}{2}}$ the electron velocity, f the frequency, and K the helical sheath impedance. If we insert into this expression the value of limiting perveance quoted above and choosing $\gamma a=1.5$, we obtain

$$CN=0.17(l/D)^{\frac{1}{2}}n^{\frac{1}{2}}.$$

[8] This assumed distribution of magnetic field can be duplicated experimentally by proper shaping of the lens pole pieces.

EQUATION OF MOTION

The electrons of the beam are considered to enter the lens from a region of zero magnetic field with uniform axial velocity $u_0=(2\eta V_0)^{\frac{1}{2}}$ and with zero radial velocity. It is sufficient to consider only the forces acting on an electron on the edge of the beam; these are, the outward radial space charge force[9]

$$F_r=-eE_r=\pi r_0^2\rho_0 e/2\pi r\epsilon_0$$

and the Lorentz forces of the magnetic field of the lens

$$F_r=-er\dot{\theta}B_z$$
$$F_\theta=e\dot{r}B_z-e\dot{z}B_r.$$

The angular velocity of an electron, emitted from a cathode at which there is no normal magnetic field, and injected into a circularly symmetric magnetic field of strength B_z on the axis, is given by Busch's theorem[10] as

$$\dot{\theta}=\eta\phi/2\pi r^2,$$

where ϕ is the total flux threading the cross section of the beam. If consideration is restricted to a beam near the axis,[11] it can be assumed that the magnetic field over the cross section of the beam is uniform, and therefore $\phi=\pi r^2B_z$. The angular velocity is then given by the equation

$$\dot{\theta}=\tfrac{1}{2}\eta B_z=\tfrac{1}{2}\omega_c. \tag{1}$$

If we make use of this relation, the equation of motion becomes

$$\ddot{r}-r(\eta B_z/2)^2=-(r/2)(\eta B_z)^2+(\pi\eta\rho_0 r_0^2/2\pi r\epsilon_0). \tag{2}$$

If we let

$$B_z=B_0/1+(z/a)^2,\quad k^2=(\eta B_0^2a^2/8V_0),$$
$$\beta_{p0}^2=\eta\rho_0/\epsilon_0u_0^2=\omega_p^2/u_0^2,$$

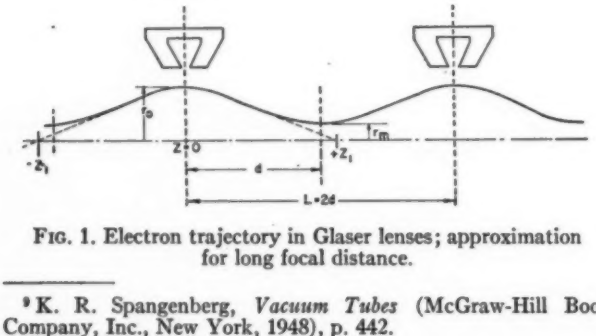


FIG. 1. Electron trajectory in Glaser lenses; approximation for long focal distance.

and $\ddot{r}=2\eta V_0(d^2r/dz^2)$, one can write Eq. (2) as

$$\frac{d^2r/r_0}{d(z/a)^2}=\frac{-k^2r/r_0}{[1+(z/a)^2]^2}+\frac{\beta_{p0}^2a^2}{2}\frac{r_0}{r}. \tag{4}$$

Here k^2 is the lens parameter, proportional to the strength of the magnetic lens, $2a$ is the half-amplitude width of the magnetic field distribution, and β_{p0} is the plasma wave number at $r=r_0$, which can be expressed in terms of the beam perveance as:

$$\beta_{p0}=(174/r_0)(I_0/V_0^{\frac{1}{2}})^{\frac{1}{2}}.$$

MAGNETIC LENS WITH LONG FOCAL DISTANCE[12]

In a focused electron beam of relatively small current (perveance less than about 10^{-7}) in which the beam diameter minimum is far from the magnetic lens, one can treat the effects of space charge and of lens focusing separately; that is, the lens is considered to provide the initial converging slope to the beam and therefore comes in only as a boundary condition. This method, of course, neglects the beam expansion in the lens itself.

The path of an electron in a Glaser lens is given by the equation[2]

$$r/a=(C_1\sin\omega\phi+C_2\cos\omega\phi)/\sin\phi \tag{5}$$

which is a solution of Eq. (4) with $\beta_p=0$, and where $\phi=\tan^{-1}a/z$, and $\omega^2=1+k^2$. In the case of an electron entering the lens parallel to the axis from $z=+\infty$, $C_2=0$, while for an electron entering along a line passing through the point $+z_1$, $C_2/C_1=\tan\omega\phi_1$ where $\phi_1=\tan^{-1}(a/z_1)$. The constant C_1 can be evaluated by noting that $r=r_0$ at $\phi=\pi/2$ $(z=0)$; thus

$$C_1=(r_0/a)/(\sin\tfrac{1}{2}\omega\pi+(C_2/C_1)\cos\tfrac{1}{2}\omega\pi).$$

The electron path is shown in Fig. 1.

The slope of an electron trajectory leaving the lens is

$$\alpha=-dr/dz=-C_1(\omega\sin\phi\cos\omega\phi-\sin\omega\phi\cos\phi)$$
$$+C_2(\omega\sin\phi\sin\omega\phi+\cos\omega\phi\cos\phi).$$

These relations are sufficient to determine the initial slope of an electron beam for given values of B_0 and V_0. The distance to the beam diameter minimum d can be found from the equation[13]

$$\frac{d}{a}=\frac{2}{\beta_{p0}a}F(R_0'), \tag{7}$$

where[14]

$$R_0'=\alpha/\beta_{p0}r_0\quad\text{and}\quad F(R_0')=e^{-R_0'^2}\int_0^{R_0'}\exp(y^2)dy.$$

[9] K. R. Spangenberg, *Vacuum Tubes* (McGraw-Hill Book Company, Inc., New York, 1948), p. 442.

[10] Reference 3, p. 36.

[11] In any circularly symmetric lens with a gap that is thin compared with the diameter of the aperture, this assumption gives a good approximation to the true field for radii out to about half the aperture radius.

[12] By focal distance is meant the distance from center of lens to the minimum in beam diameter. The conventional focal distance of a lens (distance from center of lens to point electrons cross axis) will be termed the *zero-current focal distance.*

[13] Reference 3, p. 148.

[14] This function $F(R_0')$ has been tabulated by W. Miller and A. R. Gordon, J. Phys. Chem. 35, 2785–2884 (1931).

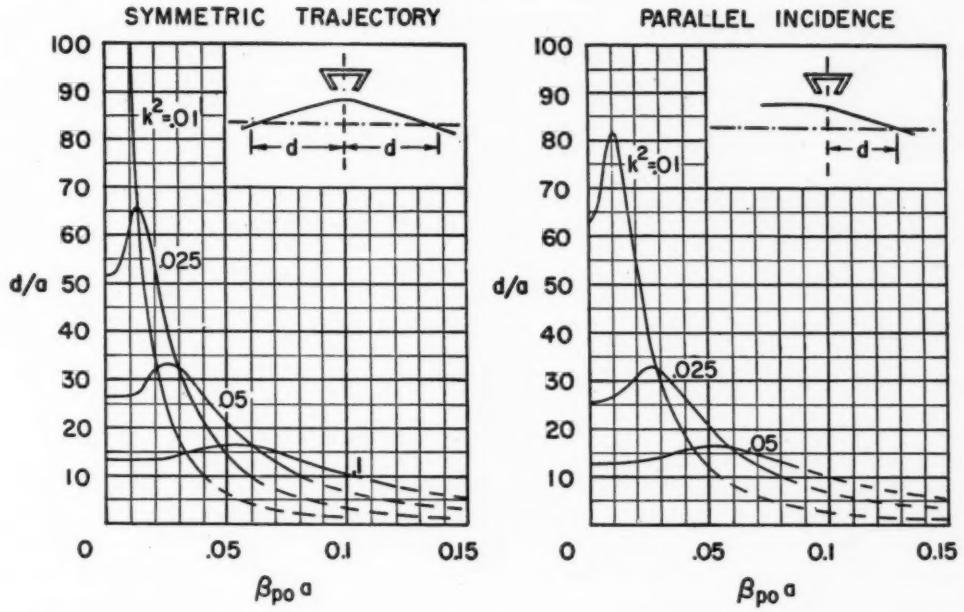


FIG. 2. Distance to beam minimum vs $\beta_{p0}a$; approximation for long focal distance.

This focal distance is plotted in Fig. 2 as a function of k^2 and $\beta_{p0}a$, where the boundary $\phi=\phi_0=\tan^{-1}(-2a/z_1)$ has been chosen as the point to match the slope of the trajectory from the lens to that of the beam. The results are insensitive to the choice of this matching point since the slope of the trajectory does not vary appreciably with z after it has been bent in toward the axis. There are, of course, other methods which can be used to match the trajectory slope in the lens to that in the beam; the one described here has been found to be satisfactory.

The beam diameter at the minimum is given by

$$r_m/r_0=\exp(-R_0'^2). \tag{8}$$

A second lens could, of course, be placed at $z=2d$ and its magnetic field strength adjusted so as to make k^2 equal to the value given in Fig. 2 for the symmetric ray.

PERIODIC MAGNETIC LENS SYSTEM

The focal distance in general decreases as the beam perveance is increased, as seen from the previous section; therefore, additional lenses must be employed to control the beam over a tube length greater than twice the focal distance. This leads one to consideration of a multiple or periodic magnetic lens system for the focusing of high-current beams.[15]

Let us consider the conditions necessary for focusing a solid electron beam by means of an iterated magnetic lens system such as that shown in Fig. 3. The envelope of the beam necessarily will be scalloped, and the period of these scallops is taken to be the period of the lens spacing. If the periodicity of the scallops and the lenses are not equal, improper focusing of the beam will occur. For example, if the conditions are such as to cause the minimum to form at a distance greater than $L/2$, at the next lens the beam will have a smaller maximum radius and will continue to contract until the increased space charge forces cause it to expand. Conversely, if the minimum is formed at a distance less than $L/2$, the beam will expand initially. Thus, any deviation from the correct focusing conditions will result in a long period variation of beam diameter superposed on the regular scallops.

We seek a solution for Eq. (4) which is symmetric about the center of the lens, considering first that the lenses are spaced far enough apart so that their fields do not overlap appreciably. It is sufficient then to consider the action of only one lens. This assumption will introduce an error in the magnitude of the magnetic force of less than 4 percent for $d/a>3$. For closer lens spacing, the effect of the nearest neighboring lens is also determined.

Symmetric solutions of Eq. (4), for certain fixed values of the parameters k^2 and $\beta_{p0}a$, were obtained by

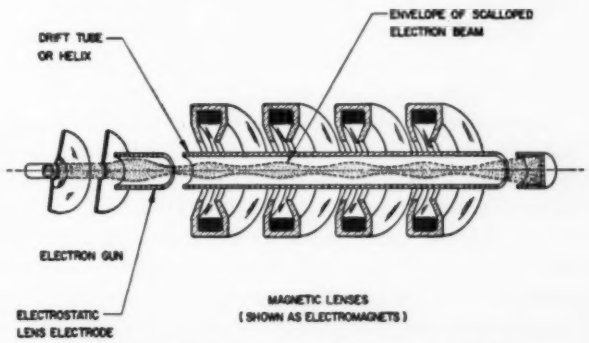


FIG. 3. Focusing of electron beam by periodic magnetic lens system.

[15] One of the earliest references to the use of two or more magnetic lenses for the control of high-current electron beams was by A. V. Haeff and L. P. Smith in U. S. Patent No. 2,225,447, filed September 13, 1939, and in a paper by A. V. Haeff and L. S. Nergaard, Proc. Inst. Radio Engrs. 28, 126–130 (1940).

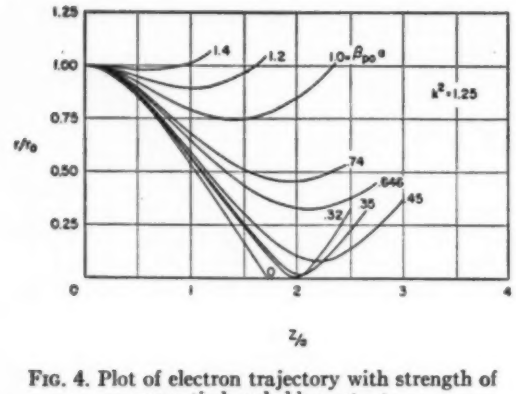


FIG. 4. Plot of electron trajectory with strength of magnetic lens held constant.

the use of an analog computer. The solution was in the form of a trajectory plot for each set of values of k^2 and $\beta_{p0}a$. Several typical trajectory plots obtained from the computer are shown in Figs. 4 and 5. The ordinate axis of these curves ($z/a=0$) represents the center of the magnetic lens and the line $r/r_0=0$ the axis of the electron beam. The envelope of the beam is generated by revolving the trajectory shown around the axis of the beam. The trajectory is symmetric about the center of the lens.

Figure 4 shows the radius as a function of distance from the lens for various values of $\beta_{p0}a$, all for the same magnet strength. These curves are seen to be qualitatively similar to those representing beams of various current densities injected into a drift tube with a given initial converging slope. The curve $\beta_{p0}a=0$ is that followed by an electron in the absence of space charge; its intersection with the axis gives the zero-current focal distance of this magnetic lens. It is seen that as the beam perveance is increased the focal distance at first increases, passes through a maximum, and finally goes to zero for a certain finite value of $\beta_{p0}a$.

Figure 5 shows the radius as a function of distance from the lens for various values of k^2 and a constant value of $\beta_{p0}a$. The focal distance increases at first with k^2 to a maximum, finally decreasing again. The behavior is very similar to that obtained by changing the initial slope of a beam entering a field-free drift tube.

The results of the solution of Eq. (4) are presented in Figs. 6 and 7. Figure 6 gives d/a, the distance to the beam minimum in units of a, as a function of $\beta_{p0}a$ for various values of k^2. For small k^2 these curves are in close agreement with the corresponding curves of Fig. 2. This graph allows one to design a periodic lens system if the beam voltage, current, and diameter are given. It is seen that for a given $\beta_{p0}a$, a number of combinations of k^2 and d/a will yield the proper focusing condition, so that some criterion for selection of the best operating point is needed. Such a criterion can be established from two considerations: (a) the depth of the scallops on the beam and (b) the chromatic aberration of the system.

Figure 7 shows the ratio of minimum to maximum beam diameter. It is desirable, of course, that the beam diameter not vary too much because of its possible effect on traveling-wave tube performance. Therefore one would choose an operating point for maximum r_m/r_0 compatible with other considerations, if the chromatic aberration is allowable in this region. The chromatic aberration will be evaluated in a subsequent section.

It is seen also from Fig. 6 that only a certain region of the chart is accessible to satisfactory focusing; that is, a combination of parameters represented by a point above the dashed line will not yield a focus of the beam. It is seen that the empirically determined line $\beta_p d=1.44$ is a reasonable approximation to the boundary of the accessible region.

EFFECT OF THE NEAREST NEIGHBORING LENS

If the perveance of the beam to be controlled with magnetic lenses is increased further (above about 2×10^{-6}), it is necessary to reduce the spacing of the lenses to such an extent that the fields overlap appreciably. Then one must consider the total field of both adjacent lenses. Considering only one nearest neighbor, we can find the beam focusing conditions between these two lenses from solutions of the modified Eq. (4),

$$\frac{d^2r/r_0}{d(z/a)^2}=-k^2\frac{r}{r_0}\left[\frac{1}{1+(z/a)^2}\pm\frac{1}{1+[(L-z)/a]^2}\right]^2+\frac{(\beta_{p0}a)^2}{2r/r_0}. \quad (9)$$

Since the magnetic force is proportional to the square of the axial field B_z, in the single lens case the sense of the fields of the adjacent magnets does not affect the analysis of beam focusing. However, if the magnets are placed closer together, so that the fields overlap, one must specify the relative direction of the alternate magnetic fields. The plus sign in the bracket in Eq. (9) refers to adjacent fields in the same direction and the minus sign to opposing fields. The practical advantage in the use of opposing magnets lies in the reduction of external fields, thereby reducing the magnet weight, for a given field, below that necessary with aiding magnets.

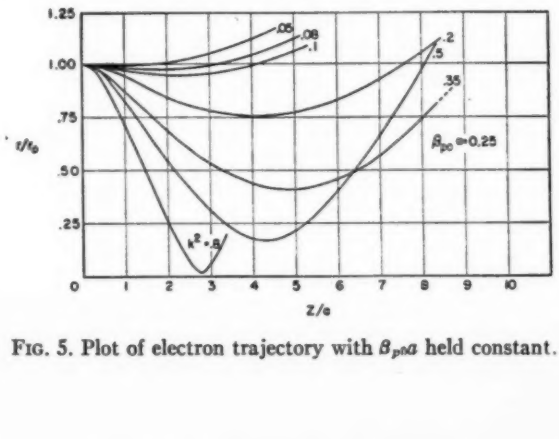


FIG. 5. Plot of electron trajectory with $\beta_{p0}a$ held constant.

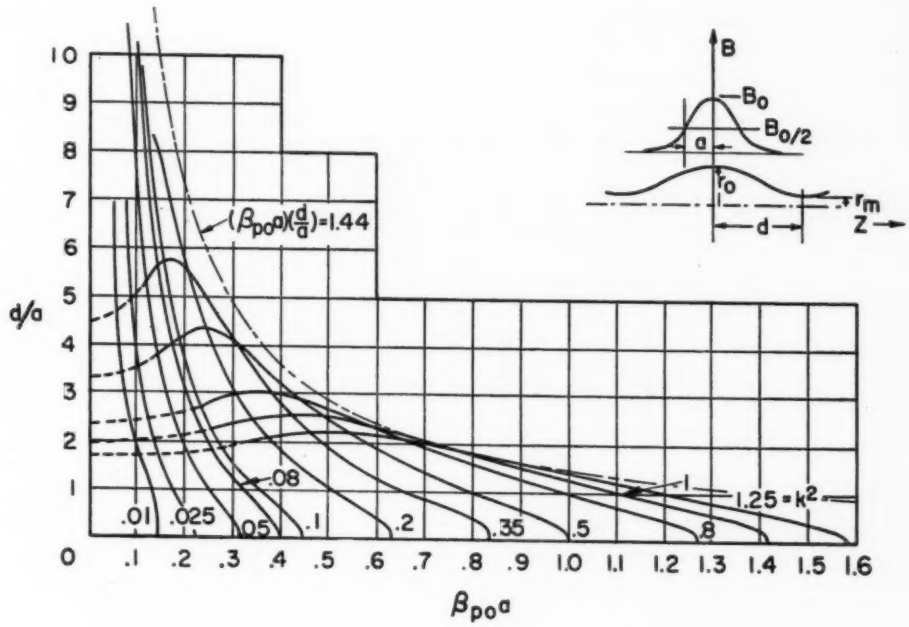


FIG. 6. Distance to beam minimum *vs* $\beta_{p0}a$; for single magnetic lens.

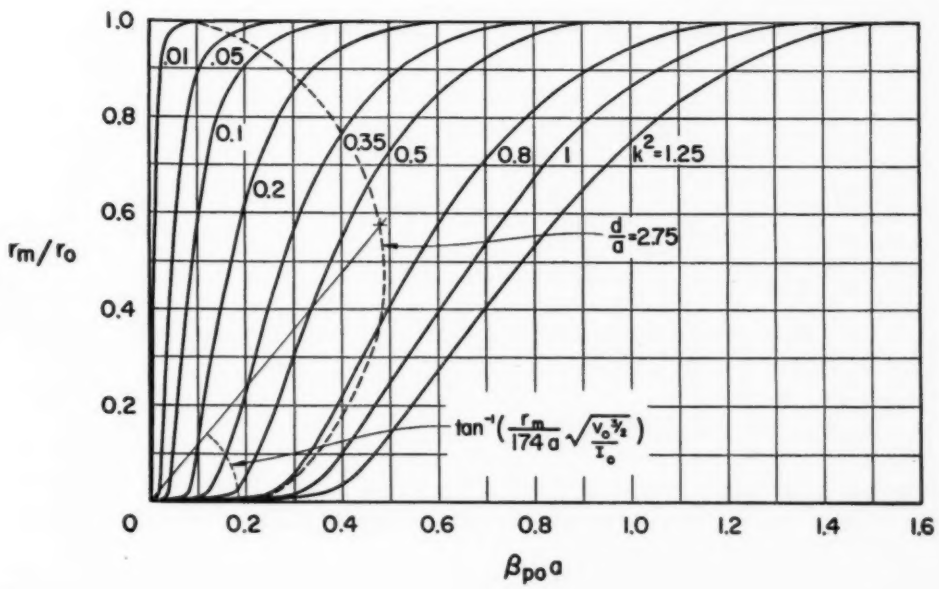


FIG. 7. Minimum beam diameter *vs* $\beta_{p0}a$; for single magnetic lens.

Equation (9) was solved for four values of $L/a = 2, 3, 4, 5.5$. The solutions are shown in Fig. 8(a) in which k^2 is plotted *vs* $\beta_{p0}a$ for $d=L/2$, and Fig. 8(b) showing the radius ratio under these conditions. Curves are given for the single lens and for double lens with aiding and opposing fields.

Notice that the curves corresponding to $d/a=2.75$ for the double lens and for the single lens agree reasonably well, while for lower values they diverge considerably. This implies that the single lens solution forms a good approximation for d/a greater than about three.

It is also seen that, owing to the double-valued nature of the curves of Fig. 8(a), regions of operation exist in which the focusing is insensitive to changes in k^2, near the vertical tangent point, and regions of insensitivity to changes in β_{p0} near the horizontal tangent point. The locus of the latter regions, which are the maxima of the curves, is drawn on the figure.

Figure 8(b) shows the radius of the minimum corresponding to the curves of Fig. 8(a). The upper part of these curves (large r_m/r_0) corresponds to the lower part (low k^2) of the curves of Fig. 8(a). It is seen that, for a given set of values of the parameters, the scallops on the beam are deeper and the limiting perveance somewhat lower in the case of opposing direction of adjacent magnets than in the aiding direction.

The values of $\beta_{p0}a$ corresponding to the vertical tangent yield the maximum perveance of the beam under the given conditions. The maximum perveance so

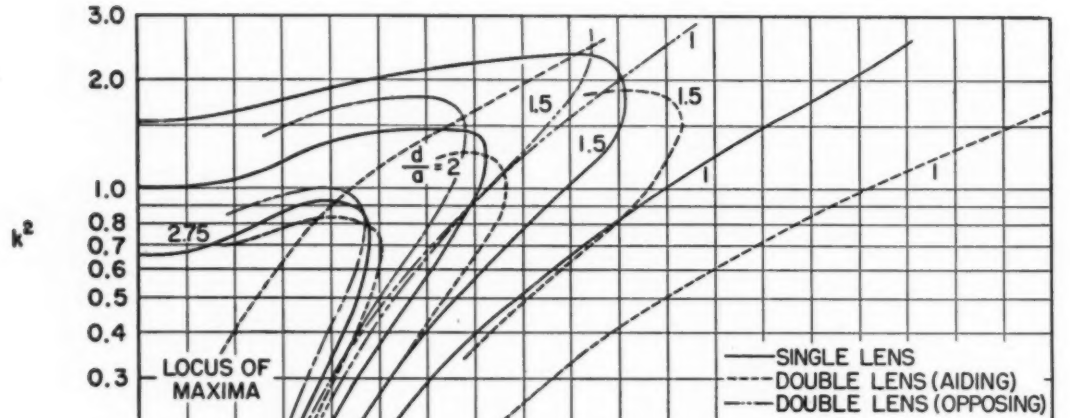


FIG. 8(a). k^2 *vs* $\beta_{po}a$ with constant distance to minimum; comparison of single and double lens solutions.

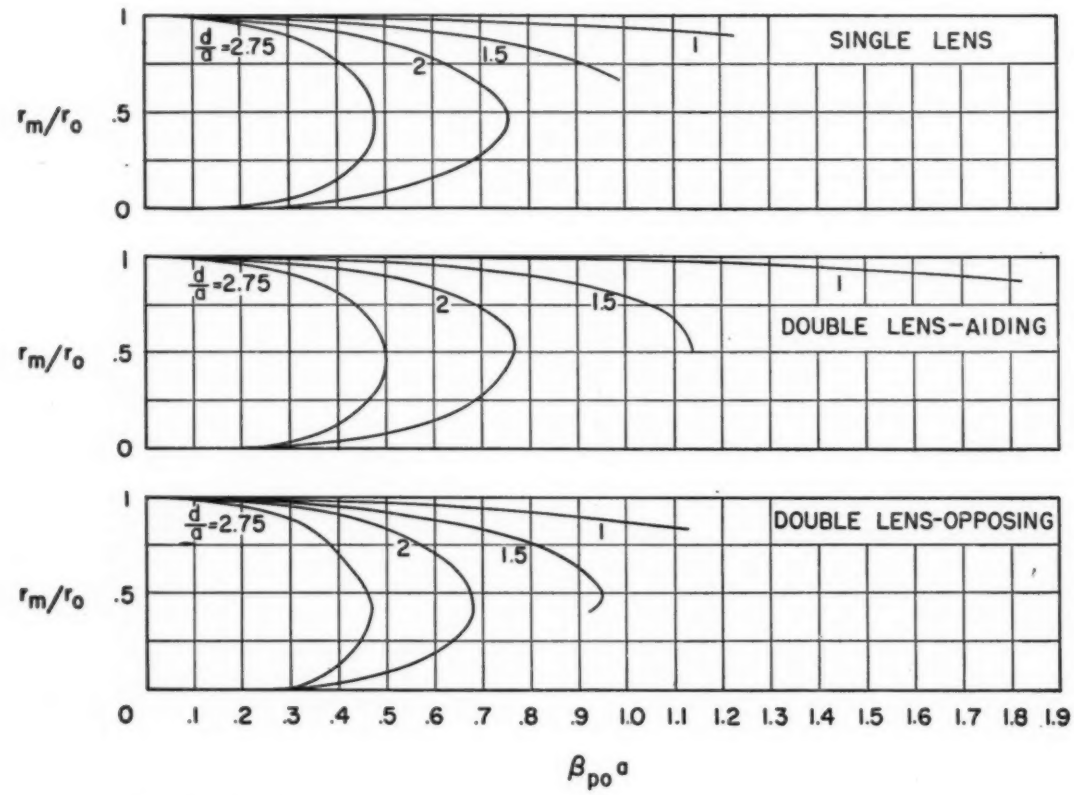


FIG. 8(b). Minimum beam diameter *vs* $\beta_{po}a$; comparison of single and double lens solutions.

determined is plotted in Fig. 9 for the single and double lens solutions. The maximum perveance $[I/V_0^{\frac{3}{2}}=38.6 \times 10^{-6}(L/D)^2]$ of a stream injected with the appropriate initial slope into a drift tube of length L and diameter D is shown for comparison. It is seen that the effect of the magnetic field, which extends into the drift tube, results in an increase in the limiting perveance, for example by a factor of two for $L/a=4$.

CHROMATIC ABERRATION

The information obtained from the solution of Eq. (4), plotted in Fig. 6, allows one to design a lens focusing system from given conditions. However, this chart does not give any indication as to the best region of operation. In choosing an operating point on the chart, relative freedom of focus conditions from fluctuations in voltage is an important consideration. The change in focal distance of a magnetic lens with voltage fluctuation is, of course, the analog of chromatic aberration in an optical lens.

An exact treatment of chromatic aberration in this periodic lens system would involve the tracing of a considerable number of rays through the entire system, using the analog computer. Since inclusion of all of the lenses of the system in the circuit analog in the computer would be rather complicated, an approximate method of calculation of this aberration will be used.

It is desired to know the variation in focal distance d with voltage V_0 as function of k^2 and $\beta_{p0}a$. Since both k^2 and β_{p0} are functions of voltage, the expression for the chromatic aberration can be written

$$\frac{d}{dV_0}(d/a)=\frac{\partial d/a}{\partial \beta_{p0}a}\frac{d\beta_{p0}a}{dV_0}+\frac{\partial d/a}{\partial k^2}\frac{dk^2}{dV_0}.$$

With insertion of the appropriate relations for $d\beta_{p0}a/dV_0$ and dk^2/dV_0, this equation can be rewritten

$$V_0\frac{d(d/a)}{dV_0}=-\frac{3}{4}\beta_{p0}a\left(\frac{\partial d/a}{\partial \beta_{p0}a}\right)_{k^2}-k^2\left(\frac{\partial d/a}{\partial k^2}\right)_{\beta_{p0}a}. \quad (10)$$

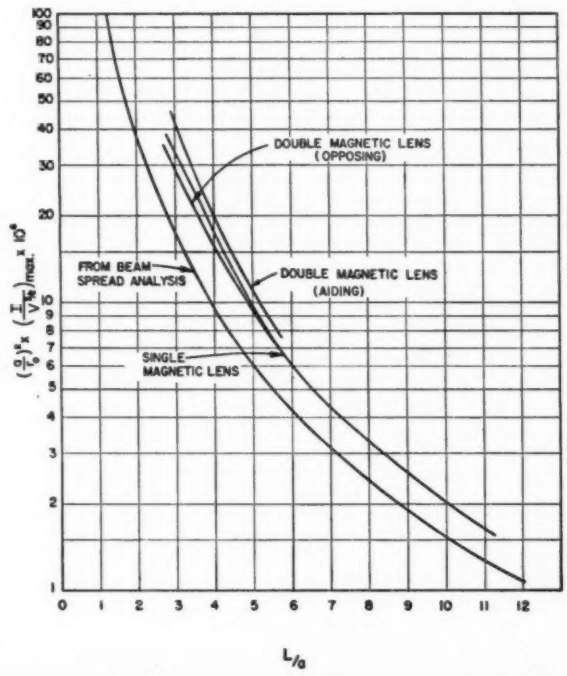


FIG. 9. Limiting perveance of lens focused electron beam *vs* length of drift tube; comparison of values with magnetic lens and from simple beam spread analysis.

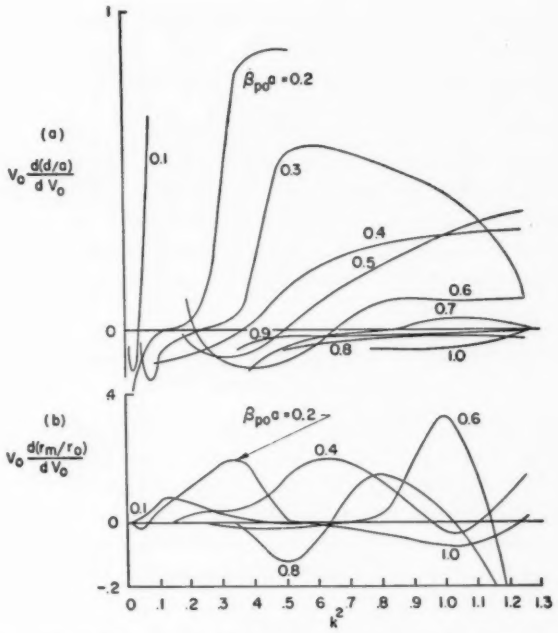


FIG. 10. A factor proportional to the chromatic aberration of a periodic magnetic lens system.

The two derivatives on the right-hand side of the equation can be evaluated graphically from Fig. 6. Equation (10) is plotted in Fig. 10(a) *vs* k^2 for various values of $\beta_{p0}a$. In the graphical evaluation of $\partial(d/a)/\partial k^2$, the data from Fig. 6 was cross plotted, and the availability of only a limited number of points reduced the accuracy with which this derivative could be evaluated; therefore, the curves of Fig. 10 should be considered only semiquantitative. It is seen that in certain regions of Fig. 6, $\partial(d/a)/\partial\beta_{p0}a$ and $\partial(d/a)/\partial k^2$ are of opposite sign, indicating that points of zero chromatic aberration exist. That is, in certain regions, the presence of space charge tends to reduce the chromatic aberration from a magnetic lens below that present in the space-charge-free case. As shown in Fig. 10, for large values of $\beta_{p0}a$ there is a wide range of values of k^2 for which the aberration is small. Examination of Fig. 7 reveals that for the values of k^2 and $\beta_{p0}a$ for which the aberration is small, the beam scallops are of relatively small magnitude.

In addition to affecting the position of the beam minimum position, fluctuations in voltage will change the depth of the minimum. If one uses a similar equation as above, one can write the change in r_m/r_0 with voltage as

$$\frac{dr_m/r_0}{dV_0}=\frac{\partial r_m/r_0}{\partial \beta_{p0}a}\frac{d\beta_{p0}a}{dV_0}+\frac{\partial r_m/r_0}{\partial k^2}\frac{dk^2}{dV_0}$$

$$V_0\frac{dr_m/r_0}{dV_0}=-\frac{3}{4}\beta_{p0}a\left(\frac{\partial r_m/r_0}{\partial \beta_{p0}a}\right)_{k^2}-k^2\left(\frac{\partial r_m/r_0}{\partial k^2}\right)_{\beta_{p0}a}. \quad (11)$$

The two derivatives can be evaluated from Fig. 7; $d(r_m/r_0)/dV_0$ is plotted in Fig. 10(b). Although this graph also should be interpreted only qualitatively, it shows that the aberration is close to zero for values of k^2 and $\beta_{p0}a$ corresponding to r_m/r_0 near unity.

These relations indicate that a periodic lens system should be designed with such values of the parameters as will yield minimum scallops of the beam, a procedure which one would be inclined to follow intuitively, from consideration of the possible effect of deep scallops on tube performance.

ENTRANCE CONDITIONS FOR PERIODIC LENS SYSTEM

The solution found for the electron trajectory in an iterated magnetic lens system is symmetric around the center of the lens and also around the line midway between the lenses; the slope of the trajectory is zero at both points. For the focusing conditions in an actual system to be as predicted by the analysis, the beam obviously must enter the magnetic field properly, that is, with the slope and radius to match the trajectory in the periodic system for the given conditions of k^2, $\beta_{p0}a$, and d/a.

This entrance condition can be met in a number of ways. Probably the simplest place to match the input trajectory to that required for stable flow is at the minimum before the first lens, which is a point of zero slope with radius and position given by Figs. 6 and 7. For example, if d/a is sufficiently large, B_z will be very small at this minimum, and the anode could be placed at this point. However, if d/a is so small that an appreciable normal magnetic field would exist at the minimum, the cathode can be moved farther away from the first magnetic lens and an electrostatic lens employed to control the beam in this transition region. If the density of the magnetic field at the cathode is denoted by B_c, the electron angular velocity is $\dot{\theta}=(\eta/2)(B_z-B_c)$, and the magnetic force term in the equation of motion becomes $(-r\eta^2/4)(B_z^2-B_c^2)$. Thus, a cathode to lens center distance of $1.5a$ will cause only ten percent error in the magnetic force. The trajectory

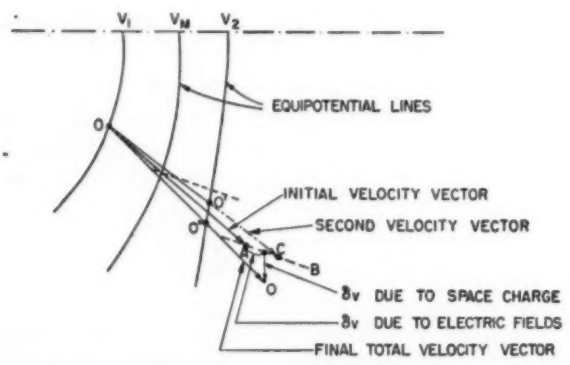


FIG. 11. Method of plotting electron trajectory including space charge.

matching could be applied at any point, of course, such as at a value of $z>d$ from the first lens (Figs. 4 and 5), using a converging beam, perhaps from a spherical cathode.

In the design of an electrostatic lens for control of the beam between the gun and the first magnetic lens, one must, of course, take the space charge expansion into account. The electrostatic lens formulas and graphs are, in general, inapplicable, and a graphical procedure appears to yield the best results. Such a simple graphical procedure for tracing an electron trajectory through a region of varying electric field, including the space charge expansion force, is outlined in the next section.

GRAPHICAL ELECTRON TRAJECTORY TRACING INCLUDING SPACE CHARGE

A number of methods for graphical tracing of electron trajectories in regions of varying electric field have been described.[16] Also, curves and formulas are available for the focal distance of an electron lens as a function of the potentials. This information is, in general, applicable only to beams of very low perveance (less than about 10^{-8}). Similar published data or procedures for high-current beams seemingly are not available.

We will describe a simple graphical procedure for tracing the trajectory of an electron in an electrostatic lens, taking account of the effect of space charge expansion of the beam. This procedure has been found to yield reasonably accurate trajectories. This method involves determination of the velocity change due to electric fields by conventional techniques, and addition to this of a radial velocity vector proportional to the space charge force.

The effect of the electric field of the lens can be found by a method due to Hansen.[17] Equipotential lines are plotted, for example, with the aid of an electrolytic tank. The vector representing the initial electron velocity at 0 is drawn proportional to $\sqrt{V_1}$ (OA, Fig. 11). A perpendicular line is erected to the mean equipotential curve V_m at the point of intersection with OA, and the line AB is drawn parallel to it. The second velocity vector OC is drawn to intersect AB and is of length proportional to $\sqrt{V_2}$. Line AC is bisected. The change in velocity of the electron due to the electric fields is thus $\frac{1}{2}AC$.

To find an approximate relation for the incremental velocity δv_{sc} due to the space-charge force, consider that

$$m\delta v_{sc}=F_r\delta t=-eE_r\delta t,$$

provided the interval δt is so short that E can be considered constant. The electric field strength due to the space charge in the beam is

$$E_r=-\rho_0 r/2\epsilon_0$$

[16] *Advances in Electronics*, edited by L. Marton (Academic Press, Inc., New York, 1950), Vol. II, which also contains an extensive bibliography of other articles.

[17] S. Hansen, Hughes Research and Development Laboratories (unpublished memorandum).

for a solid beam, and

$$E_r = -(\rho_0 r/2\epsilon_0)(1-r_i^2/r^2)$$

for a hollow beam, where r_i is the inner radius of the beam.

In time δt the electron will travel a distance $\delta s = u_m \delta t$, where $u_m = (2\eta V_m)^{\frac{1}{2}}$, so that the incremental velocity change for a solid beam is

$$\delta v_{sc} = -\frac{1}{2}\frac{\omega_p^2 r \delta s}{u_m} = \frac{\delta s I_0}{4\pi r V_m \epsilon_0},$$

where δs is the distance traveled by the electron in moving from equipotential V_1 to V_2, V_m is the mean potential between V_1 and V_2, and r is taken as the mean radius during this interval.

A trajectory plot made by this method is shown in Fig. 12. Even for the hollow low-current (1 ma) beam shown, the divergence from the space-charge free solution is seen to be significant.

When the foregoing procedure is used to arrive at the desired entrance conditions to a magnetic lens, and when the axial magnetic field is not negligible at distances greater than d from the lens, one must also include the radial Lorentz force in constructing the electron trajectory in the lens. This can be done merely by adding the term

$$F_r = -er\dot{\theta}B_z \cong -\tfrac{1}{2}er\eta B_z^2$$

to the space-charge force found previously. The magnetic field distribution is presumed known. The velocity increment is then

$$\delta v_{sc} + \delta v_{\text{mag}} = \tfrac{1}{2}\omega_p^2 r\delta s/u_m - \tfrac{1}{2}r\eta^2 B_z^2 \delta s/u_m$$
$$= \frac{r\delta s}{2u_m}(\omega_p^2 - \omega_c^2).$$

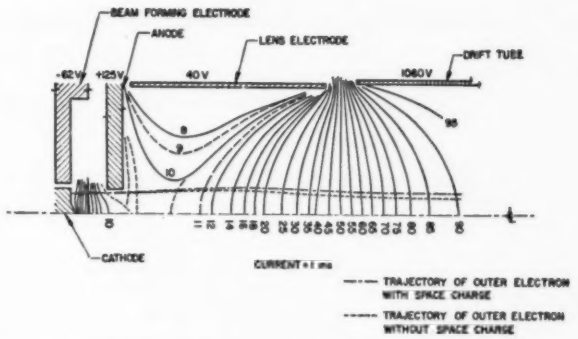


FIG. 12. Graphical plot of trajectories in a hollow electron beam; including space-charge expansion.

CONCLUSIONS

It is expected that the use of multiple magnetic-lens systems for focusing electron beams in traveling wave tubes will allow a considerable saving in weight, space, and power consumption over that required by a conventional solenoid, especially in cases where permanent magnets can be used. Thus the lens focusing technique will result in much more efficient packaging of traveling wave tubes, greatly increasing their applicability in electronic systems where space and weight are important considerations.

ACKNOWLEDGMENTS

The author wishes to acknowledge with appreciation the work of Irwin Pfeffer and Robert R. Bennett in the solution of Eqs. (4) and (9) with an analog computer, and of A. J. Donegan for other computations involved in the analysis. Thanks are also due Andrew V. Haeff and John R. Whinnery for valuable suggestions during the preparation of the manuscript.

Solution of Nonlinear Differential Equations by the Reversion Method through the First Thirteen Terms

GEORGE I. COHN,* *Illinois Institute of Technology, Chicago, Illinois*

AND

BERNARD SALTZBERG,† *Mechanics Research Department, American Machine and Foundry Company, Chicago, Illinois*

(Received July 27, 1953)

By the use of the reversion method, certain types of ordinary nonlinear differential equations can be reduced to a set of ordinary linear differential equations which may be conveniently solved by the Laplace transform. The first thirteen equations of this reduced set are tabulated in this paper.

I. INTRODUCTION

IN a recent article,[1] the authors presented a compact expression for the general ordinary linear differential equation of the reduced linear set. The ease of expanding this equation by the technique presented was illustrated by writing out in detail the 14th and 15th ordinary linear differential equations of the reduced set. These two were used because in an earlier article[2] on the subject, it was pointed out that the first thirteen were given in a paper by Van Orstrand.[3] Mr. A. C. Sim, who has made contributions to this subject,[4] has brought to our attention[5] the fact that the terms in Van Orstrand's paper were developed for reversion of algebraic formulas rather than for the solution of equations where nonpermutable operators enter. Since this is the case here, it has been requested that the terms needed to fill the gap between the 7th and 14th reversion formulas be published. Since the first seven terms are brief, they are included for convenience.

II. GENERATION OF THE ASSOCIATED SET OF LINEAR DIFFERENTIAL EQUATIONS

The reversion method may be conveniently applied to equations of the type

$$\sum_{k=1}^{n} Z_k I^k = \lambda V(t), \tag{1}$$

where $I(t)$ is the unknown response to a given driving function $V(t)$ and Z_k are functional forms of the differential and integral operators. The factor λ is introduced for the purpose of ordering the associated linear set and is set equal to unity once it has served its purpose.

A set of linear differential equations is obtained by substituting

$$I = \sum_{m=1}^{\infty} \lambda^m I_m \tag{2}$$

into (1) giving

$$\sum_{k=1}^{n} Z_k \Big(\sum_{m=1}^{\infty} \lambda^m I_m\Big)^k = \lambda V(t), \tag{3}$$

and equating coefficients of like powers of λ. The first equation in the reduced set is

$$Z_1 I_1 = V(t). \tag{4}$$

The second equation in the reduced set is

$$Z_1 I_2 = -Z_2 C_{2,2}, \tag{5}$$

where

$$C_{2,2} = I_1^2. \tag{6}$$

The third equation in the reduced set is

$$Z_1 I_3 = -Z_2 C_{3,2} - Z_3 C_{3,3}, \tag{7}$$

where

$$C_{3,2} = 2 I_2 I_1, \tag{8}$$

$$C_{3,3} = I_1^3. \tag{9}$$

The fourth equation in the reduced set is

$$Z_1 I_4 = -Z_2 C_{4,2} - Z_3 C_{4,3} - Z_4 C_{4,4}, \tag{10}$$

where

$$C_{4,2} = 2 I_3 I_1 + I_2^2, \tag{11}$$

$$C_{4,3} = 3 I_2 I_1^2, \tag{12}$$

$$C_{4,4} = I_1^4. \tag{13}$$

The fifth equation in the reduced set is

$$Z_1 I_5 = -Z_2 C_{5,2} - Z_3 C_{5,3} - Z_4 C_{5,4} - Z_5 C_{5,5}, \tag{14}$$

where

$$C_{5,2} = 2(I_4 I_1 + I_3 I_2), \tag{15}$$

$$C_{5,3} = 3(I_3 I_1^2 + I_2^2 I_1), \tag{16}$$

$$C_{5,4} = 4 I_2 I_1^3, \tag{17}$$

$$C_{5,5} = I_1^5. \tag{18}$$

The sixth equation in the reduced set is

$$Z_1 I_6 = -Z_2 C_{6,2} - Z_3 C_{6,3} - Z_4 C_{6,4} - Z_5 C_{6,5} - Z_6 C_{6,6}, \tag{19}$$

* Professor of Electrical Engineering, Illinois Institute of Technology, Chicago, Illinois.

† Research Engineer, Mechanics Research Department, American Machine and Foundry Company, Chicago, Illinois.

[1] G. I. Cohn and B. Saltzberg, J. Appl. Phys. **24**, 180 (1953).
[2] L. A. Pipes, J. Appl. Phys. **23**, 202 (1952).
[3] C. E. Van Orstrand, Phil. Mag. **19**, 366 (1910).
[4] A. C. Sim, Phil. Mag. Ser. 7, **42**, 228 (1951).
[5] A. C. Sim (private communication).

where

$$C_{6,2}=2(I_5I_1+I_4I_2+1/2I_3^2), \tag{20}$$

$$C_{6,3}=3I_4I_1^2+6I_3I_2I_1+I_2^3, \tag{21}$$

$$C_{6,4}=4I_3I_1^3+6I_2^2I_1^2, \tag{22}$$

$$C_{6,5}=5I_2I_1^4, \tag{23}$$

$$C_{6,6}=I_1^6. \tag{24}$$

The seventh equation in the reduced set is

$$Z_1I_7=-Z_2C_{7,2}-Z_3C_{7,3}-Z_4C_{7,4}-Z_5C_{7,5}-Z_6C_{7,6}-Z_7C_{7,7}, \tag{25}$$

where

$$C_{7,2}=2(I_6I_1+I_5I_2+I_4I_3), \tag{26}$$

$$C_{7,3}=3(I_5I_1^2+2I_4I_2I_1+I_3^2I_1+I_3I_2^2), \tag{27}$$

$$C_{7,4}=4(I_4I_1^3+3I_3I_2I_1^2+I_2^3I_1), \tag{28}$$

$$C_{7,5}=5(I_3I_1^4+2I_2^2I_1^3), \tag{29}$$

$$C_{7,6}=6I_2I_1^5, \tag{30}$$

$$C_{7,7}=I_1^7. \tag{31}$$

The eighth equation in the redued set is

$$Z_1I_8=Z_2C_{8,2}-Z_3C_{8,3}-Z_4C_{8,4}-Z_5C_{8,5}-Z_6C_{8,6}-Z_7C_{8,7}-Z_8C_{8,8}, \tag{32}$$

where

$$C_{8,2}=2(I_7I_1+I_6I_2+I_5I_3+I_4^2), \tag{33}$$

$$C_{8,3}=3(I_6I_1^2+2I_5I_2I_1+2I_4I_3I_1+I_4I_2^2+I_3^2I_2), \tag{34}$$

$$C_{8,4}=4I_5I_1^3+12I_4I_2I_1^2+12I_3I_2^2I_1+6I_3^2I_1^2+I_2^4, \tag{35}$$

$$C_{8,5}=5(I_4I_1^4+4I_3I_2I_1^3+2I_2^3I_1^2), \tag{36}$$

$$C_{8,6}=6I_3I_1^5+15I_2^2I_1^4, \tag{37}$$

$$C_{8,7}=7I_2I_1^6, \tag{38}$$

$$C_{8,8}=I_1^8. \tag{39}$$

The ninth equation in the reduced set is

$$Z_1I_9=-Z_2C_{9,2}-Z_3C_{9,3}-Z_4C_{9,4}-Z_5C_{9,5}-Z_6C_{9,6}-Z_7C_{9,7}-Z_8C_{9,8}-Z_9C_{9,9}, \tag{40}$$

where

$$C_{9,2}=2(I_8I_1+I_7I_2+I_6I_3+I_5I_4), \tag{41}$$

$$C_{9,3}=3(I_7I_1^2+2I_6I_2I_1+2I_5I_3I_1+I_5I_2^2+2I_4I_3I_2+I_4^2I_1+1/3I_3^3), \tag{42}$$

$$C_{9,4}=4(I_6I_1^3+3I_5I_2I_1^2+3I_4I_3I_1^2+3I_4I_2^2I_1+3I_3^2I_2I_1+I_3I_2^3), \tag{43}$$

$$C_{9,5}=5(I_5I_1^4+4I_4I_2I_1^3+2I_3^2I_1^3+6I_3I_2^2I_1^2+I_2^4I_1), \tag{44}$$

$$C_{9,6}=6I_4I_1^5+30I_3I_2I_1^4+20I_2^3I_1^3, \tag{45}$$

$$C_{9,7}=7(I_3I_1^6+3I_2^2I_1^5), \tag{46}$$

$$C_{9,8}=8I_2I_1^7, \tag{47}$$

$$C_{9,9}=I_1^9. \tag{48}$$

The tenth equation in the reduced set is

$$Z_1I_{10}=-Z_2C_{10,2}-Z_3C_{10,3}-Z_4C_{10,4}-Z_5C_{10,5}-Z_6C_{10,6}-Z_7C_{10,7}-Z_8C_{10,8}-Z_9C_{10,9}-Z_{10}C_{10,10}, \tag{49}$$

where

$$C_{10,2}=2(I_9I_1+I_8I_2+I_7I_3+I_6I_4+1/2I_5^2), \tag{50}$$

$$C_{10,3}=3(I_8I_1^2+2I_7I_2I_1+I_6I_2^2+2I_6I_3I_1+2I_5I_3I_2+2I_5I_4I_1+I_4^2I_2+I_4I_3^2), \tag{51}$$

$$C_{10,4}=4(I_7I_1^3+3I_6I_2I_1^2+3I_5I_3I_1^2+3I_5I_2^2I_1+I_4I_2^3+3/2I_4^2I_1^2+6I_4I_3I_2I_1+3/2I_3^2I_2^2+I_3^3I_1), \tag{52}$$

$$C_{10,5}=5(I_6I_1^4+4I_5I_2I_1^3+4I_4I_3I_1^3+6I_4I_2^2I_1^2+6I_3^2I_2I_1^2+4I_3I_2^3I_1+1/5I_2^5), \tag{53}$$

$$C_{10,6}=6(I_5I_1^5+5I_4I_2I_1^4+10I_3I_2^2I_1^3+5/2I_3^2I_1^4+5/2I_2^4I_1^2), \tag{54}$$

$$C_{10,7}=7(I_4I_1^6+6I_3I_2I_1^5+5I_2^3I_1^4), \tag{55}$$

$$C_{10,8}=8I_3I_1^7+28I_2^2I_1^6, \tag{56}$$

$$C_{10,9}=9I_2I_1^8, \tag{57}$$

$$C_{10,10}=I_1^{10}. \tag{58}$$

The eleventh equation in the reduced set is

$$Z_1I_{11}=-Z_2C_{11,2}-Z_3C_{11,3}-Z_4C_{11,4}-Z_5C_{11,5}-Z_6C_{11,6}-Z_7C_{11,7}-Z_8C_{11,8}-Z_9C_{11,9}-Z_{10}C_{11,10}-Z_{11}C_{11,11}, \tag{59}$$

where

$$C_{11,2}=2(I_{10}I_1+I_9I_2+I_8I_3+I_7I_4+I_6I_5), \tag{60}$$

$$C_{11,3}=3(I_9I_1^2+2I_8I_2I_1+I_7I_2^2+2I_7I_3I_1+2I_6I_4I_1+2I_6I_3I_2+I_5I_3^2+2I_5I_4I_2+I_5^2I_1+I_4^2I_3), \tag{61}$$

$$C_{11,4}=4(I_8I_1^3+3I_7I_2I_1^2+3I_6I_3I_1^2+3I_6I_2^2I_1+3I_5I_4I_1^2+I_5I_2^3+3I_3^2I_4I_1+6I_5I_3I_2I_1+3I_4I_3I_2^2+I_3^3I_2+3I_4^2I_2I_1), \tag{62}$$

$$C_{11,5}=5(I_7I_1^4+4I_6I_2I_1^3+4I_5I_3I_1^3+6I_5I_2^2I_1^2+2I_4^2I_1^3+4I_4I_2^3I_1+24I_5I_3^2+12I_4I_3I_2I_1^2+2I_3^3I_1^2+I_3I_2^4+6I_3^2I_2^2I_1), \tag{63}$$

$$C_{11,6}=6(I_6I_1^5+5I_5I_2I_1^4+10I_4I_2^2I_1^3+5I_4I_3I_1^4+10I_3^2I_2I_1^3+I_2^5I_1), \tag{64}$$

$$C_{11,7}=7(I_5I_1^6+6I_4I_2I_1^5+3I_3^2I_1^5+5I_2^4I_1^3+15I_3I_2^2I_1^4), \tag{65}$$

$$C_{11,8}=8(I_4I_1^7+7I_3I_2I_1^6+7I_2^3I_1^5), \tag{66}$$

$$C_{11,9}=9(I_3I_1^8+4I_2^2I_1^7), \tag{67}$$

$$C_{11,10}=10I_2I_1^9, \tag{68}$$

$$C_{11,11}=I_1^{11}. \tag{69}$$

The twelfth equation in the reduced set is

$$Z_1I_{12}=-Z_2C_{12,2}-Z_3C_{12,3}-Z_4C_{12,4}-Z_5C_{12,5}-Z_6C_{12,6}-Z_7C_{12,7}-Z_8C_{12,8}-Z_9C_{12,9}-Z_{10}C_{12,10}-Z_{11}C_{12,11}-Z_{12}C_{12,12}, \quad (70)$$

where

$$C_{12,2}=2(I_{11}I_1+I_{10}I_2+I_9I_3+I_8I_4+I_7I_5+1/2I_6^2), \quad (71)$$

$$C_{12,3}=3(I_{10}I_1^2+2I_9I_2I_1+I_8I_2^2+2I_8I_3I_1+2I_7I_4I_1+2I_7I_3I_2+2I_6I_5I_1+2I_6I_4I_2+I_6I_3^2+2I_5I_4I_3+I_5^2I_2+1/3I_4^3), \quad (72)$$

$$C_{12,4}=4(I_9I_1^3+3I_8I_2I_1^2+3I_7I_2^2I_1+3I_7I_3I_1^2+I_6I_2^3+3I_6I_4I_1^2+6I_6I_3I_2I_1+3I_5I_3I_2^2+3I_5I_3^2I_1+6I_5I_4I_2I_1+3/2I_5^2I_1^2+3I_4I_3^2I_2+3/2I_4^2I_2^2+3I_4^2I_3I_1+1/4I_3^4), \quad (73)$$

$$C_{12,5}=5(I_8I_1^4+4I_7I_2I_1^3+6I_6I_2^2I_1^2+4I_6I_3I_1^3+4I_5I_4I_1^3+6I_5I_3I_2I_1^2+4I_5I_2^3I_1+6I_4I_3^2I_1^2+I_4I_2^4+12I_4I_3I_2^2I_1+6I_4^2I_2I_1^2+2I_3^2I_2^3+4I_3^3I_2I_1), \quad (74)$$

$$C_{12,6}=6(I_7I_1^5+5I_6I_2I_1^4+5I_5I_3I_1^4+10I_5I_2^2I_1^3+5/2I_4^2I_1^4+10I_4I_2^3I_1^2+20I_4I_3I_2I_1^3+10/3I_3^3I_1^3+15I_3^2I_2^2I_1^2+5I_3I_2^4I_1+1/6I_2^6), \quad (75)$$

$$C_{12,7}=7(I_6I_1^6+6I_5I_2I_1^5+6I_4I_3I_1^5+15I_4I_2^2I_1^4+15I_3^2I_2I_1^4+20I_3I_2^3I_1^3+3I_2^5I_1^2), \quad (76)$$

$$C_{12,8}=8(I_5I_1^7+7I_4I_2I_1^6+7/2I_3^2I_1^6+21I_3I_2^2I_1^5+35/4I_2^4I_1^4), \quad (77)$$

$$C_{12,9}=9I_4I_1^8+72I_3I_2I_1^7+84I_2^3I_1^6, \quad (78)$$

$$C_{12,10}=10I_3I_1^9+45I_2^2I_1^8, \quad (79)$$

$$C_{12,11}=11I_2I_1^{10}, \quad (80)$$

$$C_{12,12}=I_1^{12}. \quad (81)$$

The thirteenth equation in the reduced set is

$$Z_1I_{13}=-Z_2C_{13,2}-Z_3C_{13,3}-Z_4C_{13,4}-Z_5C_{13,5}-Z_6C_{13,6}-Z_7C_{13,7}-Z_8C_{13,8}-Z_9C_{13,9}-Z_{10}C_{13,10}-Z_{11}C_{13,11}-Z_{12}C_{13,12}-Z_{13}C_{13,13}, \quad (82)$$

where

$$C_{13,2}=2(I_{12}I_1+I_{11}I_2+I_{10}I_3+I_9I_4+I_8I_5+I_7I_6), \quad (83)$$

$$C_{13,3}=3(I_{11}I_1^2+2I_{10}I_2I_1+I_9I_2^2+2I_9I_3I_1+2I_8I_3I_2+2I_8I_4I_1+2I_7I_5I_1+2I_7I_4I_2+I_7I_3^2+2I_6I_4I_3+2I_6I_5I_2+I_6^2I_1+I_5I_4^2+I_5^2I_3), \quad (84)$$

$$C_{13,4}=4(I_{10}I_1^3+3I_9I_2I_1^2+3I_8I_3I_1^2+3I_8I_2^2I_1+3I_7I_4I_1^2+I_7I_2^3+6I_7I_3I_2I_1+3I_6I_5I_1^2+6I_6I_4I_2I_1+3I_6I_3I_2^2+3I_6I_3^2I_1+3I_5I_4I_2^2+6I_5I_4I_3I_1+3I_5I_3^2I_2+3I_5I_4I_2^2+3I_5^2I_2I_1+I_4I_3^3+I_4^3I_1), \quad (85)$$

$$C_{13,5}=5(I_9I_1^4+4I_8I_2I_1^3+4I_7I_3I_1^3+6I_7I_2^2I_1^2+4I_6I_4I_1^3+6I_6I_3I_2I_1^2+4I_6I_2^3I_1+2I_5^2I_1^3+6I_5I_3^2I_1^2+12I_5I_4I_2I_1^2+I_5I_2^4+12I_5I_3I_2^2I_1+6I_4^2I_3I_1^2+6I_4^2I_2^2I_1+12I_4I_3^2I_2I_1+4I_4I_2I_3^2+2I_3^3I_2^2+I_3^4I_1), \quad (86)$$

$$C_{13,6}=6(I_8I_1^5+5I_7I_2I_1^4+5I_6I_3I_1^4+10I_6I_2^2I_1^3+5I_5I_4I_1^4+20I_5I_3I_2I_1^3+10I_5I_2^3I_1^2+10I_4I_3^2I_1^3+30I_4I_3I_2^2I_1^2+5I_4I_2^4I_1+10I_4^2I_2I_1^3+I_3I_2^5+10I_3^2I_2^3I_1+10I_3^3I_2I_1^2), \quad (87)$$

$$C_{13,7}=7(I_7I_1^6+6I_6I_2I_1^5+15I_5I_2^2I_1^4+6I_5I_3I_1^5+3I_4^2I_1^5+30I_4I_3I_2I_1^4+10/3I_4I_2^3I_1^3+15I_3I_2^4I_1^2+5I_3^3I_1^4+30I_3^2I_2^2I_1^3+I_2^6I_1), \quad (88)$$

$$C_{13,8}=8(I_6I_1^7+7I_5I_2I_1^6+7I_4I_3I_1^6+21I_4I_2^2I_1^5+21I_3^2I_2I_1^5+14I_2^5I_1^3+35I_3I_2^3I_1^4), \quad (89)$$

$$C_{13,9}=9(I_5I_1^8+8I_4I_2I_1^7+4I_3^2I_1^7+28I_3I_2^2I_1^6+14I_2^4I_1^5), \quad (90)$$

$$C_{13,10}=10(I_4I_1^9+9I_3I_2I_1^8+12I_2^3I_1^7), \quad (91)$$

$$C_{13,11}=11(I_3I_1^{10}+5I_2^2I_1^9), \quad (92)$$

$$C_{13,12}=12I_2I_1^{11}, \quad (93)$$

$$C_{13,13}=I_1^{13}. \quad (94)$$

JOURNAL OF APPLIED PHYSICS VOLUME 25, NUMBER 2 FEBRUARY, 1954

Distortion of a Crystal by Point Imperfections*

J. D. ESHELBY
Department of Physics, University of Illinois, Urbana, Illinois
(Received August 17, 1953)

The expression $\mathbf{u}=c\mathbf{r}/r^3$, (where c is a constant) sometimes assumed for the displacement around a point imperfection (interstitial or substitutional impurity, lattice vacancy) gives a nonzero stress at the surface of the solid. The additional "image displacement" necessary to insure that this stress vanishes is usually neglected, but may be important. For example, it accounts for from 30 to 50 percent of the volume change produced by such defects. This and other effects of the image term are discussed. Miller and Russel have pointed out that for a point imperfection near the center of a sphere the apparent volume change deduced from measurements of the x-ray lattice constant is greater than the geometrical volume change. It is shown that the reverse is true when the defect is near the surface, and that for a large number of defects scattered uniformly through the sphere the geometrical and x-ray expansions are equal. It can be shown quite generally that a body of arbitrary shape is expanded uniformly by a statistically uniform distribution of point imperfections, and that the x-ray diffraction pattern is altered in the way to be expected for such an expansion. To establish this, however, it is essential to take the image terms into account.

I. INTRODUCTION

As a model for the distortion of a crystal lattice by a substitutional or interstitial atom or a vacant lattice site, a number of authors have taken a center of dilatation in an isotropic elastic continuum. In many cases calculations have been made using the elastic displacement appropriate to a center of dilatation in an infinite medium. This displacement cannot be correct since it would give a nonvanishing stress at the free surface of the body. It is convenient to regard the actual displacement as the sum of the displacement in an infinite medium and an "image" displacement due to the perturbing effect of the boundary.

It is the object of this paper to show that the image term cannot be neglected in certain applications. In particular, if it is omitted the volume change due to one or more point singularities will be underestimated by a factor of 1.5 for metals or 1.8 for alkali halides. Again, if a calculation is made in which the effects of the free boundary are ignored, it is found that any nonspherical body uniformly filled with point singularities would suffer a change of shape as well as a change of size. When the boundary effect is taken into account there is a change of volume without change of shape. Moreover, a calculation of the change of x-ray lattice constant gives a result consistent with the geometrical deformation only if the image displacements are included. Thus, unusually, the boundary effects simplify the problem instead of complicating it.

II. A POINT SINGULARITY IN AN INFINITE BODY

The elastic displacement caused by a center of dilatation in an infinite medium is[1]

$$\mathbf{u}^\infty = c\frac{\mathbf{r}}{r^3} = -c\,\mathrm{grad}(1/r), \qquad (1)$$

where c is a constant, the "strength" of the singularity. (We use the affix ∞ to emphasize that (1) is only valid in an infinite medium.) The displacement has the same form as the field of an electrostatic point charge. The real justification for taking (1) as a rough description of the elastic field of a point imperfection in an infinite crystal which has been idealized as a homogeneous isotropic elastic continuum is that it is the only spherically symmetrical displacement which satisfies the equations of elasticity and does not increase with r. Still, it is convenient to have some sort of detailed elastic model. As a model of an interstitial or substitutional atom we might take an elastic sphere of radius $(1+\epsilon)r_0$ forced into a spherical hole of radius r_0 in an infinite block of the same material. It is easy to show that for $r>r_0$ the displacement is given by (1) with $c=\epsilon r_0^3(1+\sigma)/3(1-\sigma)$ and that for $r<r_0$ there is a uniform compression. (σ is Poisson's ratio.) The surface of the hole suffers an outward displacement c/r_0^2, increasing the volume within it by $4\pi c$.

This model must not be taken too literally. The lattice constant of gold is decreased by a little dissolved silver. The misfitting sphere model would suggest that therefore a little dissolved gold would increase the lattice constant of silver, and this is not true. All that we can hope for is that a particular type of singularity in a particular matrix will be characterized by a constant c which can be used consistently to describe various phenomena.

Again, a foreign atom in a lattice exerts forces on its neighbors differing from the "standard" forces they would experience in the perfect lattice. When we assimilate the lattice to a continuum, the standard forces are absorbed in the elastic properties of the medium, but the additional forces due to the imperfection are still outstanding. Hence as an alternative to the sphere-in-hole model we may take a cluster of point forces to represent a point lattice imperfection in the elastic approximation. For a sufficiency symmetrical relation of the interstitial or substitutional

* This work was supported by the U. S. Office of Naval Research.
[1] A. H. Love, *Mathematical Theory of Elasticity* (Cambridge University Press, Cambridge, England, 1924).

atom and the lattice, the forces will be equivalent to three equal "double forces without moment" at right angles. The displacement will then be of the form (1) which can be considered[1] as due to a distribution of body force

$$\mathbf{F}(\mathbf{r}) = -12\pi cK \operatorname{grad}\delta(\mathbf{r}). \tag{2}$$

(K is the bulk modulus.) In a less symmetrical case (e.g., carbon in iron) the cluster of forces will be equivalent to three unequal double forces without moment and the resulting displacement will be more complicated than (1). However, it is probably not justifiable to introduce this refinement without at the same time considering the anisotropy of the material, and we shall not consider it further.

The displacement (1) can be produced by heating a point of the elastic medium, assumed to be nonconducting. This can be seen by imagining that the misfit in the first model was produced by heating a sphere with originally no misfit, or in terms of the second model by noticing that the thermal stress[2] due to a temperature distribution $T(\mathbf{r})$ is the same as the stress produced by a density of body force proportional to grad T, which for the hot spot $T=\text{const}\,\delta(\mathbf{r})$ would be of the form (2). This analogy will be useful in discussing a body containing a large number of imperfections.

In general the stress associated with a displacement $\mathbf{u}$ whose Cartesian components are u_1, u_2, u_3 is

$$p_{ij} = \lambda\delta_{ij}\operatorname{div}\mathbf{u} + \mu\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right), \tag{3}$$

where λ and μ are Lamé's constants. From the electrostatic interpretation it is at once clear that the divergence and curl of (1) vanish, so that the stress produced by the singularity when in an infinite body is simply

$$p_{ij}^{\infty} = 2\mu\frac{\partial u_i^{\infty}}{\partial x_j}. \tag{4}$$

III. A POINT SINGULARITY IN A FINITE BODY WITH A STRESS-FREE SURFACE

Consider now a center of dilatation in a finite body with a free surface. First mark out the surface S of the proposed body in an infinite block of material and introduce the singularity. The displacement is given correctly by (1). Across any surface element of S there is a stress† $p_{ij}n_j dS$, where $\mathbf{n}$ is the normal to S. Thus if we remove the material outside S the displacement will continue to be $\mathbf{u}^{\infty}$ only if we apply a distribution of surface traction $p_{ij}n_j$ to S. Removing this distribution to give a body with a stress-free surface is equivalent to applying an additional distribution $-p_{ij}n_j$. The elastic state will then be given by

$$\mathbf{u} = \mathbf{u}^{\infty} + \mathbf{u}^{I}, \quad p_{ij} = p_{ij}^{\infty} + p_{ij}^{I}, \tag{5}$$

where the image stress p_{ij}^{I} is the stress which surface tractions $-p_{ij}n_j$ would produce in the body and the image displacement $\mathbf{u}^{I}$ is related to p_{ij}^{I} by (3). p_{ij}^{I} is clearly free of singularities within S and satisfies

$$(p_{ij}^{I} + p_{ij}^{\infty})n_j = 0 \quad \text{on} \quad S. \tag{6}$$

Unlike $\mathbf{u}^{\infty}$, $\mathbf{u}^{I}$ has in general a nonvanishing divergence.

If the surface traction and surface displacement are prescribed (they are not, of course, independent) the elastic field inside the body can be found by integration. When only the surface traction is given, we need the appropriate elastic Green's function, known for only a few simple shapes. Since we know the image traction $p_{ij}^{I}n_j$ but not $\mathbf{u}^{I}$ on S we cannot in general calculate the image field in the body.

The change in volume of the solid can be divided into parts ΔV^{∞} and ΔV^{I} arising from the two terms in (5). We have at once

$$\Delta V^{\infty} = \int_S \mathbf{u}^{\infty}\cdot\mathbf{n}\,dS = 4\pi c, \tag{7}$$

the integral being, according to (1), c times the solid angle subtended by the surface at the singularity. This result can be seen at once for the sphere-in-hole model. When the sphere is inserted, the volume of the hole increases by $4\pi c$, and since $\operatorname{div}\mathbf{u}^{\infty}=0$ outside the hole, this increase is transmitted unchanged to S. Again for the ideal mathematical singularity for which (1) holds for all $\mathbf{r}$ we have strictly not $\operatorname{div}\mathbf{u}^{\infty}=0$ but rather

$$\operatorname{div}\mathbf{u}^{\infty} = -c\nabla^2(1/r) = 4\pi c\delta(\mathbf{r}) \tag{8}$$

with a delta-function of expansion at $r=0$. A formal volume integration gives (7).

Although we cannot find the image deformation in detail, we can find ΔV^{I} with the help of the rule that the volume change of a body subjected to a distribution of surface traction $\mathbf{T}$ per unit area is

$$\frac{1}{3K}\int_S \mathbf{r}\cdot\mathbf{T}\,dS = \frac{1}{3K}\int_S x_i p_{ij} n_j\,dS, \tag{9}$$

where p_{ij} is the stress produced by $\mathbf{T}$ and K is the bulk modulus. For

$$\int x_i p_{ij} n_j\,dS = \int \frac{\partial}{\partial x_j}(x_i p_{ij})\,dv = \int p_{jj}\,dv = 3K\int \operatorname{div}\mathbf{u}\,dv \tag{9'}$$

since $\partial x_i/\partial x_j = \delta_{ij}$ and in elastic equilibrium $\partial p_{ij}/\partial x_j = 0$.

To find ΔV^{I} we must put $p_{ij}n_j = p_{ij}^{I}n_j = -p_{ij}^{\infty}n_j$ with p_{ij}^{∞} from (4). But the operator $x_i(\partial/\partial x_i)$ applied to $u_k^{\infty} = cx_k/r^3$ merely multiplies it by -2. Hence the

[1] S. Timoshenko, *Theory of Elasticity* (McGraw-Hill Book Company, Inc., New York, New York, 1934).

† Throughout the paper we use the convention that a repeated suffix is to be summed over the values 1, 2, 3.

integral (9) is equal to the integral (7) times $4\mu/3K$ and so

$$\Delta V^I = 4\pi c \frac{2(1-2\sigma)}{1+\sigma}. \tag{10}$$

Thus the total volume change on introducing a center of dilatation of strength c into any homogeneous isotropic body with a stress-free surface is

$$\Delta V = \Delta V^\infty + \Delta V^I = \gamma \Delta V^\infty = 4\pi c\gamma, \tag{11}$$

where the constant

$$\gamma = 3(1-\sigma)/(1+\sigma) \tag{12}$$

is about 1.5 for metals ($\sigma \sim \frac{1}{3}$) and 1.8 for alkali halides ($\sigma \sim \frac{1}{4}$). This result has already been given by Seitz.[3] It is also implicit in the known result for the interaction energy of a center of dilatation and an external hydrostatic pressure.[4] This energy is the product of ΔV and the hydrostatic pressure.

If in (9′) we take p_{ij} to be any state of purely internal stress, so that the surface integral is zero we arrive at the known result that the volume of a body is unaffected by internal stresses provided Hooke's law is valid. The volume expansion we have calculated is not in disagreement with this. Consider the sphere-in-hole model of Sec. II, and suppose for clearness that ϵ is negative. Then the unstrained state is a body with a spherical hole of radius r_0 containing a sphere of radius $r_0(1-|\epsilon|)$ and an unoccupied volume $4\pi r_0^3|\epsilon| = 4\pi|c|\gamma$. In the strained state the surfaces of the hole and sphere have been drawn together and welded. Since the total volume of *material* is unchanged, the volume enclosed by the outer surface must have decreased by an amount equal to the volume eliminated between sphere and hole. This simple derivation does not supersede our more elaborate treatment since to relate the expansion to other effects arising from the stress field of the imperfections we need to know something about $\mathbf{u}^\infty$ and $\mathbf{u}^I$ separately. If there are departures from Hooke's law, the volume change of the material is not quite zero. This is discussed in Sec. VI.

IV. DEFORMATION PRODUCED BY A LARGE NUMBER OF SINGULARITIES

If there are a number of singularities in the body, we have

$$\mathbf{u}^\infty(\mathbf{r}) = \sum_n c_n \frac{\mathbf{r}-\mathbf{r}_n}{|\mathbf{r}-\mathbf{r}_n|^3}, \tag{13}$$

with summation over all imperfections. The image stress $p_{ij}{}^I$ is that stress which has no singularities within S and on S annuls the surface traction calculated from (13). $\mathbf{u}^I$ is the associated displacement.

If the defects are all alike we have

$$\Delta V/V = 4\pi\gamma cf/\Omega, \tag{13′}$$

where f is the atomic fraction of defects and Ω is the volume per atom of the matrix.

Of special interest is the case where there are a large number of identical centers spread through the body, with an approximately uniform density of n centers per unit volume. Here we can also say something about the change of shape if we allow ourselves to replace the actual distribution by a continuous distribution of infinitesimal centers with the same total strength per unit volume. (We shall try to justify this below.) If as before the body is embedded in an infinite matrix this averaged displacement is clearly

$$\langle \mathbf{u}^\infty(\mathbf{r}) \rangle = cn \int_V \frac{\mathbf{r}-\mathbf{r}'}{|\mathbf{r}-\mathbf{r}'|^3} dv(\mathbf{r}') \tag{14}$$

at any rate outside S. (It is not at once clear that we can give any meaning to the average displacement in a medium riddled with singularities.)

The comparison already made with thermal expansion suggests that a body uniformly filled with infinitesimal centers of dilatation will, like a uniformly heated solid, expand uniformly. The following is a formal proof. We first show that the displacement given by the volume integral (14) can be dupilicated outside S by a distribution of body force over S of uniform magnitude

$$p = K\Delta V/V = 4\pi cn\gamma K$$

per unit area and directed along the outward normal at each point. A point force $\mathbf{F}$ in an infinite medium produces a displacement[1]

$$\mathbf{u} = A\mathbf{F}/r + B\mathbf{r}(\mathbf{r}\cdot\mathbf{F})/r^3,$$
$$B = 1/16\pi\mu(1-\sigma), \quad A = (3-4\sigma)B$$

at a vector distance $\mathbf{r}$ from it. Each element of the surface contributes a force $d\mathbf{F} = p\mathbf{n}dS$ and the net displacement is

$$u_i = p\int_S \left\{ \frac{A}{r}\delta_{ij} + \frac{B}{r^3}x_i x_j \right\} n_j dS$$
$$= p\int \frac{\partial}{\partial x_j}\{\ \}dv = p(A-B)\int \frac{-x_i}{r^3}dv$$
$$= \langle u_i^\infty \rangle.$$

(We have taken the point of observation to be the origin.)

Suppose now that S is marked out in an infinite medium, that the uniform distribution if infinitesimal centers of dilatation is introduced and that a layer of body force equal and opposite to that just described is distributed over S. The combined effect of these last two steps gives zero displacement outside S and leaves S its original shape and size. If the unstrained matrix is now cut away, the elastic state of the body bounded by S is unchanged, but the layer of body force becomes

[3] Frederick Seitz, Revs. Modern Phys. **18**, 384 (1946).
[4] See, for example, J. D. Eshelby, Trans. Roy. Soc. (London) **A244**, 87 (1951).

a hydrostatic pressure p acting on the surface. When this is removed S undergoes the uniform expansion

$$\langle u_i \rangle = (4/3)\pi n c \gamma x_i. \tag{15}$$

Though $\langle \mathbf{u}^\infty \rangle + \mathbf{u}^I$ represents a uniform expansion, $\langle \mathbf{u}^\infty \rangle$ and $\mathbf{u}^I$ separately do not, except for a spherical body. Thus omission of the image terms in addition to giving a volume change lacking the factor γ (which could be absorbed in the usually unknown constant c) would also give a nonuniform expansion. Suppose, for example, that S is an ellipsoid with semi-axes $\mathfrak{a}$, $\mathfrak{b}$, $\mathfrak{c}$. Then $\langle \mathbf{u}^\infty \rangle = -cn \operatorname{grad} \varphi$ where

$$\varphi = \tfrac{1}{2}(D - Ax_1 - Bx_2 - Cx_3) \tag{16}$$

with the usual notation for the potential of an ellipsoid.[5] Thus if image terms are neglected, the ellipsoid becomes another ellipsoid with semi-axes $\mathfrak{a}+\Delta\mathfrak{a}$, etc., where

$$\frac{\Delta\mathfrak{a}}{\mathfrak{a}} = cnA, \quad \frac{\Delta\mathfrak{b}}{\mathfrak{b}} = cnB, \quad \frac{\Delta\mathfrak{c}}{\mathfrak{c}} = cnC. \tag{17}$$

Since $A+B+C=4\pi$ the volume change agrees with (7). As an example, if $\mathfrak{a}:\mathfrak{b}:\mathfrak{c}=2:1:1$ then also $A:B:C=2:1:1$ very nearly. Again, a cube would become barrel-shaped, with the increase in distance between opposite face centers equal to about three times the increase in distance between two adjacent corners. These departures from uniform expansion, if they existed, would be hard to detect by macroscopic measurements, but as we shall see, there is a similar effect on the lattice constants determined by x-ray diffraction, and this would be easily observable.

We must now try to justify replacing the sum (13) by the integral (14). Just this question arises when we replace the field of a set of gravitating or electrified particles by the field of a continuous body. This kind of problem can be attacked in various ways, none completely satisfying; the following seems a plausible line for the elastic case.

The observable macroscopic displacement at a point of a body in which the displacement fluctuates on a microscopic scale is essentially the displacement of the centroid of a region large compared with the scale of the fluctuations. It is therefore reasonable to define the gross (or macroscopic) displacement at a point as the average of the actual (microscopic) displacement taken over the volume of a sphere whose radius R is large enough to contain many fluctuations, but is small compared with the dimensions of the body.

Let us find the gross displacement $\langle \mathbf{u}^\infty \rangle$ corresponding to (14) for a point inside the body. The volume integral over the sphere R of the displacement caused by a singularity at $\mathbf{r}$ relative to its center is equal to the attraction on a charge c at $\mathbf{r}$ produced by a distribution of charge of unit density filling the sphere, i.e., $(4/3)\pi c\mathbf{r}$ if $r<R$ and $(4/3)\pi cR^3\mathbf{r}/r^3$ if $r>R$. Thus

$$\langle \mathbf{u}^\infty \rangle = \frac{c}{R^3} \sum_{r_n<R} \mathbf{r}_n + c \sum_{r_n>R} \frac{\mathbf{r}_n}{r_n^3}.$$

As R increases and becomes reasonably large compared with $n^{-\frac{1}{3}}$ the first term should be small, being proportional to the position vector of the centroid of a large number of points taken at random in a sphere. At the same time, as R increases it becomes more and more reasonable to replace the second term by an integral since the distance of all the singularities from the point at which their effects are summed is large compared with the mean distance between them. We thus reach (14) but with a sphere of radius R about the point of observation excluded from the integration. However, by a known result of potential theory (14) is unaltered by this omission provided the sphere lies wholly within S.

Having defined the gross displacement we can calculate a gross stress from it with the aid of (3). This should be the true macroscopic stress, i.e., it should give correctly the surface traction required to prevent relative motion of the faces of a macroscopic cut made in the material.

We can obviously generalize these results to a distribution of imperfections in which n is a function of position. The microscopic $\mathbf{u}^\infty$ is still given by (13) but the gross displacement is

$$\langle \mathbf{u}^\infty \rangle = c \int_V \frac{n(\mathbf{r}')(\mathbf{r}-\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|^3} dv(\mathbf{r}'), \tag{18}$$

and in the averaging process we must have $n^{-\frac{1}{3}} \ll R \ll n^{-\frac{1}{3}} |\operatorname{grad}(n^{-\frac{1}{3}})|$.

The gross or microscopic $p_{ij}{}^I$ is the stress produced by surface tractions $-p_{ij}{}^\infty n_j$ calculated from (13) or (14,18). These two surface tractions will differ by rapidly fluctuating quantities with a "wavelength" of the order of $n^{-\frac{1}{3}}$ whose effect will be confined to a surface layer of the same order. Thus in the bulk of the material the image quantities are efficiently smoothed by St. Venant's principle whether we smooth the infinity quantities or not, and we need not distinguish between their gross and microscopic values.

V. EFFECT OF POINT SINGULARITIES ON THE X-RAY DIFFRACTION PATTERN

Miller and Russel[6] originally suggested that there is a difference between the geometrical volume change of a crystal containing imperfections and the volume change deduced from the change of x-ray lattice constant. They calculate that for a uniform distribution of point imperfections the latter should be about twice the former.[7] Huang[8] and Teltow[9] found that they were

[5] O. D. Kellog, *Foundations of Potential Theory* (Julius Springer, Berlin, 1929), p. 194.

[6] P. H. Miller, Jr., and B. R. Russel, J. Appl. Phys. **23**, 1163 (1952).

[7] They have since withdrawn this factor. P. H. Miller, Jr., and B. R. Russel, J. Appl. Phys. **24**, 1248 (1953).

[8] K. Huang, Proc. Roy. Soc. (London) **A190**, 102 (1947).

[9] J. Teltow, Ann. Physik **12**, 111 (1953).

equal. These authors omit image effects. Miller and Russel calculate the image effect for a single imperfection at the center of a sphere. They take a collection of such spheres with the interstices filled up to represent a body containing many imperfections and argue that the image effect is negligible because the surface traction across these spheres is by no means zero. As we have seen, the effect of the image term on the volume change can be taken into account quite generally, and we shall find that the same is true for its effect on the x-ray lattice constant.

For an undisturbed crystal the scattering power plotted in reciprocal space will be a set of patches with their maxima at the points of the reciprocal lattice. If the crystal is distorted by external forces or internal imperfections the patches will be displaced and deformed. We may regard the new maxima as defining a new reciprocal lattice and we can then speak of the deformation of the reciprocal lattice corresponding to a deformation of the crystal.

Let the position of a point of the crystal or reciprocal lattice before and after distortion be

$$\mathbf{r}=L_i\mathbf{a}_i, \quad \mathbf{r}+\Delta\mathbf{r}=(L_i+\Delta L_i)\mathbf{a}_i$$

and

$$\mathbf{B}=h_i\mathbf{b}_i, \quad \mathbf{B}+\Delta\mathbf{B}=(h_i+\Delta h_i)\mathbf{b}_i.$$

Miller and Russel derive the following relation valid within a region around the origin of reciprocal space which decreases as the degree of crystal distortion increases:

$$\Delta h_i A_{ij}=-h_i B_{ij} \quad (j=1,2,3), \tag{19}$$

where

$$A_{ij}=\sum L_iL_j, \quad B_{ij}=\sum \Delta L_iL_j,$$

$\sum$ implying summation over all points of the crystal lattice. We have simplified equations (1,2,3) in reference 6, appendix 1, by taking the origin at the lattice point nearest the center of gravity of the crystal: then the sums $\sum L_i$ will differ from zero by integers small compared with the number of atoms in the crystal and may be neglected. Equation (19) has the solution

$$\Delta h_i=-h_kC_{ki}, \quad C_{ki}=B_{kj}A_{ji}' \tag{20}$$

where A_{ij}' is the matrix reciprocal to A_{ij}. C_{ij} is not in general symmetric even if B_{ij} is. Thus (20) is an infinitesimal affine transformation in reciprocal space. In other words the reciprocal lattice points are shifted as if they were embedded in an imaginary elastic continuum which undergoes a deformation in which all the strain and rotation components

$$e_{ij}=\tfrac{1}{2}(\partial\Delta h_i/\partial h_j+\partial\Delta h_j/\partial h_i)=-\tfrac{1}{2}(C_{ij}+C_{ji}),$$
$$\omega_{ij}=\tfrac{1}{2}(\partial\Delta h_i/\partial h_j-\partial\Delta h_j/\partial h_i)=-\tfrac{1}{2}(C_{ij}-C_{ji})$$

are in general different from zero and constant for small h_i. This implies, for example, that from x-ray measurements of the position of low-order spots alone, a distorted cubic crystal could not be distinguished from an unstrained monoclinic crystal slightly misoriented.

Following Huang and Teltow we may replace the displacement expressed as a sum of the effects of the separate imperfections by an integral if the distribution is statistically uniform. This corresponds to the averaging process of the previous section, and we may put‡ $a\Delta L_i=\langle u_i\rangle$. The summation $\sum$ may also be replaced by an integral and we have

$$A_{ij}=\frac{1}{a^5}\int_V x_ix_j dv, \quad B_{ij}=\frac{1}{a^5}\int_V x_i\langle u_j\rangle dv. \tag{21}$$

But we know that when the image terms are included $\langle u_i\rangle$ has the form (15), so that $A_{ij}=(4/3)\pi ncB_{ij}$ and (20) has the solution

$$\frac{\Delta h_1}{h_1}=\frac{\Delta h_2}{h_2}=\frac{\Delta h_3}{h_3}=-(4/3)\pi nc\gamma.$$

(The A_{ij} are products of inertia of the crystal and $|A_{ij}|$ could only vanish if the moment of inertia of the crystal about some axis were zero.) Thus the reciprocal lattice contracts uniformly in the way we should expect for a crystal which had suffered the uniform expansion (15).

This simple result would be destroyed if the image terms were omitted. We should have

$$B_{ij}=-\frac{cn}{a^5}\int_V x_i\frac{\partial\varphi}{\partial x_j}dv,$$

where

$$\varphi=\int_V dv/r$$

is the potential of the specimen if filled with charge of unit density. For an ellipsoidal crystal φ is a quadratic function of the coordinates (Eq. (16)), and it follows that $-\Delta h_1/h_1$, $-\Delta h_2/h_2$, $-\Delta h_3/h_3$ are equal, respectively, to the three quantities (17). In this case the (incorrect) deformation of the reciprocal lattice agrees with the (incorrect) macroscopic deformation with image effects omitted. For a body of arbitrary shape this is no longer true, since there is no simple relation between A_{ij} and B_{ij}. X-ray measurements would indicate that the cubic unit cell had become monoclinic, and its change of volume would bear no simple relation to the change in volume of the crystals calculated with or without image effects. These complications are implicit in Huang's calculations. Each of the two terms in his expression (I) is similar to our (14). Thus Huang's (I) does not represent a uniform expansion except for a sphere, and even for the sphere the expansion lacks a factor γ. Actually Huang is concerned with the more general question of the distribution of scattering power between the reciprocal lattice points. It is easy to see that in his expressions (II) and (III) c must be replaced by $c\gamma$. Since c is eventually eliminated with the help of his

‡ For simplicity we now suppose the crystal is simple cubic with lattice spacing a.

Eq. (18) in which also c must be replaced by $c\gamma$ the final result is unaffected, as also is the important result that the sharpness of the diffraction peaks is unaltered by the uniform distribution of defects.

Miller and Russel found by calculating the $\sum L_i \Delta L_i$ numerically that for a single singularity at the center of a sphere the expansion deduced from x-rays was 2.5 times the geometrical expansion. They originally deduced from this that there should be a similar difference (with the factor reduced from 2.5 to about 2) when there are a large number of imperfections distributed at random.

The previous discussion indicates that this is incorrect, but does not show exactly how the discrepancy arises. This discrepancy exists whether image terms are introduced or not (though introducing them would reduce the factor from 2 to $1+1/\gamma$) and for simplicity we omit them in the following discussion.

Let us find the effect of a single imperfection at a point ξ relative to the center of a spherical crystal of radius R. We can save calculation by finding the dilatation of the reciprocal lattice and not its detailed distortion. Putting $\langle u_i \rangle = c(x_i - \xi_i)/|\mathbf{x}-\xi|^3$ in (21) we find after some manipulation

$$\frac{\Delta V^\infty}{V} = -\frac{\partial \Delta h_i}{\partial h_i} = \frac{15}{4\pi}\frac{c}{R^5}\left(1+\xi_i\frac{\partial}{\partial \xi_i}\right)\int_V \frac{dv}{|\mathbf{x}-\xi|}.$$

The integral is the gravitational potential at ξ in a solid sphere of unit density and has the value $\frac{2}{3}\pi(3R^2-\xi^2)$. Thus

$$\Delta V^\infty = 10\pi c(1-\xi^2/R^2). \tag{22}$$

If we exclude from the integration a sphere of radius r_0 about ξ a term $-r_0^2/R^2$ must be introduced inside the bracket. Since r_0 can presumably be taken of atomic dimensions we shall omit it.

The reason for the discrepancy now becomes clear. The expression (22) is greater or less than the geometrical ΔV^∞ according as ξ is greater or less than $\sqrt{(3/5)}R = 0.77R$. For $\xi = 0$ it is 2.5 times the geometrical value, in agreement with Miller and Russel. If there are a large number of defects scattered at random throughout the sphere, the average contribution to ΔV^∞ per imperfection is given by replacing ξ^2 in (22) by its mean value for all points of the sphere,

$$\langle \xi^2 \rangle = 4\pi \int_0^R r^4 dr \Big/ 4\pi \int_0^R r^2 dr = \tfrac{3}{5}R^2.$$

This gives a mean x-ray ΔV^∞ of $4\pi c$, in agreement with the geometrical value (7). There seems to be nothing objectionable in the replacement of sums by integrals, and presumably Miller and Russel's summation procedure would give the geometrical value if carried out for a reasonable number of imperfections scattered throughout a sphere.

As a corollary we see that with a given number of defects the x-ray expansion depends strongly on their (nonuniform) distribution, while the geometrical expansion is, of course, independent of the distribution. Thus, for a sphere the ratio of the x-ray ΔV^∞ to the geometrical ΔV^∞ can be given any value between 2.5 and 0 by sweeping the defects towards the center or surface. When the two expansions are corrected for the image terms the ratio lies between $1+1.5/\gamma$ and $1-1/\gamma$.

VI. DISCUSSION

The equality of the x-ray and geometrical expansions of a crystal uniformly filled with point imperfections has a bearing on the question whether they are of Schottky or Frenkel type. Revised calculations by Miller and Russel[7] gave a difference of not more than about 10 or 20 percent between the two expansions in place of their original factor of two. The residual difference may be partly the result of the neglect of image terms.

Turning to the factor γ introduced into the expansion by the image terms it looks at first sight as if one could redefine the strength of an imperfection as $c\gamma$ and then forget about the image effects. This is not so. The contributions to the expansion from $\mathbf{u}^\infty$ and $\mathbf{u}^I$ are in the ratio $1:(\gamma-1)$. But the ratio of their contributions to any other phenomenon will in general be quite different, since $\mathbf{u}^\infty$ is a rapidly fluctuating function of position associated with a pure shear, whereas $\mathbf{u}^I$ is smooth except near the surface with an associated stress which is chiefly a hydrostatic pressure in a body of reasonable shape; c is thus the basic constant which we might hope to determine from measurements of one effect and apply in calculating the value of another. The following examples illustrate this.

Nabarro[10] has discussed the hardness of dilute alloys. The hardening is due almost entirely to p_{ij}^∞, the slowly varying p_{ij}^I playing scarcely any part. He eliminates c with the help of the observed change of lattice constant with composition but neglects the image effect. This can be put right by dividing his theoretical constant $\alpha = 0.05$ by γ before comparing it with the values deduced from experiment. The spread of the latter is too great to say if this gives any improvement.

Dexter[11] has calculated the effects of $\mathbf{u}^\infty$ and $\mathbf{u}^I$ on the electrical conductivity of metals. The former gives a temperature independent and the latter a temperature dependent contribution. They are of the same order at room temperature.

Overhauser[12] has considered the modification of rates of diffusion by the strain due to point imperfections. The effects arising from $\mathbf{u}^\infty$ and $\mathbf{u}^I$ depend differently on temperature, and the image term may become important at high temperatures.

Zener[13] has calculated the change of shear modulus caused by the $\mathbf{u}^\infty$ field of point imperfections. Evidently

[10] F. R. N. Nabarro, Proc. Phys. Soc. (London) **58**, 669 (1946).
[11] D. L. Dexter, Phys. Rev. **87**, 768 (1952).
[12] A. W. Overhauser, Phys. Rev. **90**, 393 (1953).
[13] C. Zener, Acta Cryst. **2**, 163 (1949).

the dilatation due to the image field should give rise to an analogous change of bulk modulus. (As we have seen, $\mathbf{u}^\infty$ produces no dilatation of the material between the imperfections.) In fact we should have a change

$$\frac{\Delta K}{K}=\frac{d\log K}{d\log V}\frac{\Delta V^I}{V}=\frac{d\log K}{d\log V}\frac{\gamma-1}{3\gamma}\frac{dl}{l} \tag{23}$$

proportional to the relative change dl/l of lattice constant but of opposite sign, since $d\log K/d\log V$ is negative. According to a simple theory[14] verified quite well by experiment $-d\log K/d\log V$ is equal to $2\Gamma+4/3$, where Γ is Grüneisen's constant. Zener's $\Delta\mu/\mu$ on the other hand depends on the shear strain energy associated with an imperfection and so is proportional to $|dl/l|$.

By an atomic calculation Dienes[15] has found that one percent of interstitials increase the bulk modulus of copper by 6.8 percent and of sodium by 1.9 percent. His calculations deal in principle with what happens in the immediate neighborhood of the imperfection, so it is reasonable to suppose that they take account of $\mathbf{u}^\infty$ and that the effects of $\mathbf{u}^I$ should be added. The results of Dienes and of Huntington and Seitz[16] suggest that we may take $4\pi c/\Omega$ to be about 0.6 and 0.9, respectively, for an interstitial in copper or sodium. By (13′) and (23) the corresponding percentage changes for one percent interstitials are lattice constant 0.9 (Cu), 1.4 (Na), bulk modulus -4.6 (Cu), -6.9 (Na). Our calculation cannot claim any accuracy, but it suggests that the image term is important, perhaps even large enough to change the sign of $\Delta K/K$.

The change of shear modulus found by Zener's argument depends on nonlinear behavior in the region close to the imperfection (see below) and it is therefore in principle already included in a calculation such as Dienes' which considers the balance of interatomic forces near the imperfection.

It was pointed out at the end of Sec. III that for a purely elastic model the actual volume of material was unaltered. Zener[17] has shown how in such a case the departure from Hooke's law gives a volume change, and we ought to consider what effect this will have on our calculations. This change has the value

$$\Delta V'=-(1+d\log K/d\log V)W_d/K \\ -(1+d\log\mu/d\log V)W_s/K,$$

where W_d and W_s are the total dilatational and shear strain energies associated with the internal stress. The first term can be considered as a small correction to ΔV^I corresponding to using in (9) the value of K appropriate to the lattice as expanded. The shear strain associated with the sphere-in-hole model of Sec. II is $8\pi\mu c^2/r_0^3$ (see reference 8). If we take the case of interstitials in a face-centered cubic material and put r_0 equal to the distance from the interstitial to the nearest face-centered atom we have

$$\Delta V'/V=-(1+d\log\mu/d\log V)16\pi(\mu/K)(c/\Omega)^2 f.$$

This is about one tenth of (13′) for interstitials in copper if as before we take $c/\Omega\sim 0.05$. However, even if the ratio were quite large our previous calculations would not be upset. Of the shear strain energy $8\pi\mu c/r_0^3$ associated with an imperfection nearly ninety percent is contained within a sphere of radius $2r_0$. Thus the correction comes entirely from an additional expansion confined to the immediate neighborhood of the imperfection. This extra volume change is transmitted to the surface by the displacement $\mathbf{u}^\infty$ which must have the form (1) as soon as the elastic region is reached. Thus the effect we are considering merely "renormalizes" the constant c in (1) and it is correct to use the modified value not only in (13′) but also in discussing any other phenomenon in which only strains in the elastic region are important.

Point imperfections may expand the lattice by mechanisms other than misfit. Take for example an alkali halide crystal with doubly charged foreign cations in some positions together with an equal number of vacant anion sites. The continuum model is a dielectric with positive and negative charges embedded in it. Round each there is an elastic stress field due to electrostriction. The dilatation is large near a charge and falls off rapidly with distance. The stress field is essentially that of a center of dilatation. A rough calculation suggests that its strength is of the same order as for a reasonable degree of misfit. Since it is proportional to the square of the charge, the effect should exist although the number of positive and negative charges is equal. Again, if atoms of different valency are introduced substitutionally into a metal, the resulting change in the number of electrons in the conduction band will produce a change of lattice constant quite apart from any misfit effects.

It is not the aim of this paper to advocate the elastic approach, but to show that when it is used neglect of the presence of the free surface may lead to qualitative untidinesses (such as nonuniform deformation by a uniform distribution of defects) and quantitative errors which may be important.

[14] J. C. Slater, Phys. Rev. **57**, 744 (1940).
[15] G. J. Dienes, Phys. Rev. **86**, 228 (1952).
[16] H. B. Huntington and F. Seitz, Phys. Rev. **61**, 315 (1942).
[17] C. Zener, Trans. Am. Inst. Mining Met. Engrs. **147**, 361 (1942).

JOURNAL OF APPLIED PHYSICS VOLUME 25, NUMBER 2 FEBRUARY, 1954

An Approximate Method of Evaluating Integral Transforms*

ARMEN H. ZEMANIAN†
New York University, New York, New York

(Received July 22, 1953)

A new approximate means of evaluating integral transforms is developed. The process, which is very simple and quite general, has been applied to the Fourier, Laplace, Mellin, and Hankel transforms. The method may also be successful as a numerical means of evaluating an integral which has a highly oscillatory term in its integrand.

1. INTRODUCTION

THE importance of integral transformations in the various mathematical sciences needs hardly be mentioned. So far as electrical engineering is concerned, network theory has been so completely based on the Fourier and Laplace transforms that the frequency response of systems had, for a while, assumed a much greater importance than its time response to the engineer. Furthermore, the various other transformations such as the Mellin and Hankel transforms are applicable in solving network problems for time-varying systems.[1] These transforms also find use in various electromagnetic field and wave problems, to mention just another application. However it is not uncommon that a particular transform cannot be evaluated due to its complexity and so a simple general method of obtaining an approximate expression for the transformation would be quite useful. The proposed method accomplishes this and it may also be successful as a numerical means of evaluating an integral which has a highly oscillatory term in its integrand.

Although a considerable number of methods[2–12] exist for the evaluation of the Fourier and Laplace transforms, no simple method has been found in the literature for the evaluation of any integral transform. Furthermore, it is felt that the proposed method is as simple as any of the existing methods for the evaluation of the Fourier and Laplace transforms.

2. STATEMENT OF THE GENERAL PROCEDURE

The general procedure will first be outlined before it is developed and illustrated by some examples.

(a) The function which is to be transformed is approximated by a set of confluent straight-line segments, none of which have an infinite slope.

(b) The changes in slope between straight-line segments are calculated.

(c) These values are substituted into the general expression for the approximation of the transformation and the resulting expression evaluated.

3. GENERAL DEVELOPMENT OF FORMULAS

Consider the following general transformation.

$$F(s)=\int_a^b f(t)H(s,t)dt, \tag{1}$$

where t is a real variable, s may be either real or complex and the limits may assume any real values. The case where the limits become infinite will be considered subsequently. It is assumed that the following indefinite integrals can be found:

$$G(s,t)=\int H(s,t)dt,$$

$$K(s,t)=\int G(s,t)dt.$$

Integrating expression (1) by parts, the following is obtained:

$$F(s)=[f(t)G(s,t)]_a^b-\int_a^b \frac{df}{dt}G(s,t)dt. \tag{2}$$

This is the basis of an approximation method in which the interval (a,b) is divided into $n+1$ sections and the function $f(t)$ approximated in each section by a straight line segment. The average slope of the function in each section is approximated by the slope of the corresponding segment and these segments are taken to join end on end. In short, the graph of $f(t)$ is approximated by a polygon whose vertices need not lie on this graph.

* This paper is based on a portion of a thesis which has been accepted by the faculty of the Graduate Division of the College of Engineering of New York University in partial fulfillment of the requirements for the degree of Doctor of Engineering Science.

† Instructor, Electrical Engineering Department, New York University.

[1] J. A. Aseltine, "Transforms for Linear Time-Varying Systems," University of California, Los Angeles, Report 52-1 (February, 1952).

[2] A. V. Bedford and G. L. Fredendall, Proc. Inst. Radio Engrs. **30**, 440–57 (1942).

[3] G. S. Brown and D. P. Campbell, *Principles of Servo-mechanisms* (John Wiley and Sons Inc., New York, 1948), first edition, Chapter 11.

[4] W. J. Cunningham, J. Appl. Phys. **18**, 656–664 (1947).

[5] C. H. Dawson, "Approximation of Transient Response from Frequency Response Data," Am. Inst. Elec. Engrs. Technical paper 53-248, presented at Am. Inst. Elec. Engrs. Summer General Meeting, Atlantic City, New Jersey, June 19, 1953.

[6] A. Erdelyi, Phil. Mag. **34**, 533–537 (1943).

[7] L. N. G. Filon, Proc. Roy. Soc. Edinburgh **49**, 38–47 (1928–29).

[8] T. Roddam, Wireless World **58**, 292–295 (1952).

[9] J. B. Russell, J. Math. and Phys. **12**, 274–290 (1932).

[10] H. A. Samulon, Proc. Inst. Radio Engrs. **39**, 175–86 (1951).

[11] C. J. Tranter, *Integral Transforms in Mathematical Physics* (John Wiley and Sons Inc., New York, 1951), first edition.

[12] H. A. Wheeler, Proc. Inst. Radio Engrs. **25**, 359–85 (1939).

Now let S_{fk} be the slope of the kth straight line segment in the approximation, t_k the abscissa of the point of intersection between the kth and $(k+1)$st straight line segments, $t_0=a$ and $t_{n+1}=b$. By using S_{fk} in place of df/dt in the previous expression and rearranging terms, the following approximation to $F(s)$ is obtained:

$$F_A(s)=[f(t)G(s,t)]_a^b+S_{f1}K(s,a)$$
$$-S_{f(n+1)}K(s,b)+\sum_{k=1}^{n}K(s,t_k)\Delta S_{fk}, \quad (3)$$

where $\Delta S_{fk}=S_{f(k+1)}-S_{fk}$. This is the desired result.

The transformations considered below have infinite limits. If in addition to the customary assumption, assuring the existence of $F(s)$ in expression (1), the assumption is made that the first term on the right-hand side of expression (2) remains bounded at the infinite limit, then the resulting approximation remains valid when the limits of integration become infinite. Furthermore, the straight line approximation must be such that the transform of the approximating function exists.

In general, the transformation, $F(s)$, is a complex function and the inverse transform is of the form

$$f(t)=\int_C P(s,t)F(s)ds,$$

where the path of integration, C, in the s plane is suitably chosen insuring, among other things, that the integral converges. To apply the development to this inverse transform it will first be necessary to break the integral up into its real and imaginary parts, if possible, so that it is the sum of two real integrals. Then the above procedure may be applied to each part to obtain the total approximate expression for the inverse transform.

4. DETERMINATION OF A BOUND ON THE ERROR

An upper bound on the magnitude of the error between the actual and approximate transformations can be obtained in the following manner. Consider again the general transformation (1). The error in the s domain produced by the approximation in the t domain, $f_A(t)$, is given by the following:

$$\mathcal{E}(s)=\int_a^b [f(t)-f_A(t)]H(s,t)dt, \quad (5)$$

so that

$$|\mathcal{E}(s)|\leq\int_a^b |f(t)-f_A(t)|\,|H(s,t)|\,dt. \quad (6)$$

Furthermore for $a\leq t\leq b$ and s suitably restricted, the magnitude of the kernel, $H(s,t)$, may be bounded. That is,

$$|H(s,t)|\leq M,$$

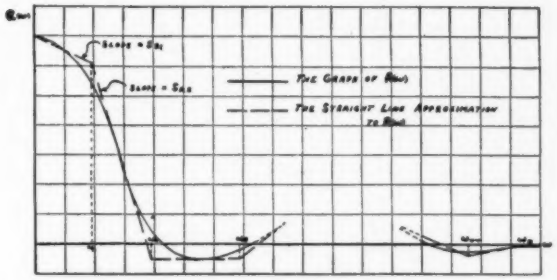


FIG. 1. Illustration of straight line approximation of $R(\omega)$.

where M is a positive, real number. Thus,

$$|\mathcal{E}(s)|\leq M\int_a^b |f(t)-f_A(t)|\,dt. \quad (7)$$

This integral can be easily evaluated by calculating the area between the graphs of $f(t)$ and $f_A(t)$, however this upper bound on the magnitude of the error will in general be considerably greater than the actual error.

5. EXAMPLE: THE CONSTRUCTION OF THE APPROXIMATE FOURIER COSINE TRANSFORM

As an example of the method of constructing the approximate expression for a transformation, the following Fourier cosine transform will be considered:

$$W(t)=\frac{2}{\pi}\int_0^\infty R(\omega)\cos\omega t d\omega. \quad (8)$$

As applied to network theory, $W(t)$ and $R(\omega)$ may represent the response to an impulse applied at $t=0$ and the real part of the corresponding system function, respectively. Approximating $R(\omega)$ by n straight-line segments as shown in Fig. 1 such that the $(n+1)$st segment coincides with the horizontal axis beyond t_n and formally applying the development as outlined before, the approximate transform is obtained

$$W_A(t)=-\frac{2}{\pi t^2}(S_{R1}+\sum_{k=1}^{n}\Delta S_{Rk}\cos\omega_k t). \quad (9)$$

It can be seen that this expression approaches the indeterminate form of zero over zero as t approaches zero. However, an expression for this value may be obtained by integrating the exact expression for the initial value by parts and applying the approximation as before. This yields

$$W_A(0)=\frac{1}{\pi}\sum_{k=1}^{n}\omega_k^2\Delta S_{Rk}.$$

Finally the expression for the bound on the error is

$$|\mathcal{E}(t)|=|W(t)-W_A(t)|$$
$$\leq\frac{2}{\pi}\int_0^\infty |R(\omega)-R_A(\omega)|\,d\omega. \quad (10)$$

TABLE I.

Transform	Exact transform	Approximate transform	Initial value	Error bound
I. Fourier transform	a. $F(j\omega)=\int_{-\infty}^{+\infty} f(t)e^{-j\omega t}dt$	$F_A(j\omega)=-\frac{1}{\omega^2}\sum_{k=1}^{n}\Delta S_{fk}e^{-j\omega t_k}$	$F_A(0)=\frac{1}{2}\sum_{k=1}^{n}\Delta S_{fk}t_k^2$	$\lvert\varepsilon(j\omega)\rvert\leq\int_{-\infty}^{+\infty}\lvert f(t)-f_A(t)\rvert dt$
	b. $f(t)=\frac{1}{2\pi}\int_{-\infty}^{+\infty}F(j\omega)e^{j\omega t}d\omega$	$f_A(t)=-\frac{1}{\pi t^2}(S_{R1}+\sum_{k=1}^{n}\Delta S_{Rk}\cos\omega_k t)+\frac{1}{\pi t^2}\sum_{k=1}^{m}\Delta S_{Ik}\sin\omega_k t$	$f_A(0)=\frac{1}{2\pi}\sum_{k=1}^{n}\Delta S_{Rk}\omega_k^2$	$\lvert\varepsilon(t)\rvert\leq\frac{1}{2\pi}\int_{-\infty}^{+\infty}\lvert R(\omega)-R_A(\omega)\rvert d\omega+\frac{1}{2\pi}\int_{-\infty}^{+\infty}\lvert I(\omega)-I_A(\omega)\rvert d\omega$
	c. $R(\omega)=\int_0^{\infty}W(t)\cos\omega t dt$	$R_A(\omega)=-\frac{1}{\omega^2}(S_{W1}+\sum_{k=1}^{n}\Delta S_{Wk}\cos\omega t_k)$	$R_A(0)=\frac{1}{2}\sum_{k=1}^{n}\Delta S_{Wk}t_k^2$	$\lvert\varepsilon(\omega)\rvert\leq\int_0^{\infty}\lvert W(t)-W_A(t)\rvert dt$
	d. $W(t)=\frac{2}{\pi}\int_0^{\infty}R(\omega)\cos\omega t d\omega$	$W_A(t)=-\frac{2}{\pi t^2}(S_{R1}+\sum_{k=1}^{n}\Delta S_{Rk}\cos\omega_k t)$	$W_A(0)=\frac{1}{\pi}\sum_{k=1}^{n}\Delta S_{Rk}\omega_k^2$	$\lvert\varepsilon(t)\rvert\leq\frac{2}{\pi}\int_0^{\infty}\lvert R(\omega)-R_A(\omega)\rvert d\omega$
	e. $I(\omega)=-\int_0^{\infty}W(t)\sin\omega t dt$	$I_A(\omega)=-\frac{W(0)}{\omega}+\frac{1}{\omega^2}\sum_{k=1}^{n}\Delta S_{Wk}\sin\omega t_k$		$\lvert\varepsilon(\omega)\rvert\leq\int_0^{\infty}\lvert W(t)-W_A(t)\rvert dt$
	f. $W(t)=-\frac{2}{\pi}\int_0^{\infty}I(\omega)\sin\omega t d\omega$	$W_A(t)=\frac{2}{\pi t^2}\sum_{k=1}^{n}\Delta S_{Ik}\sin\omega_k t$		$\lvert\varepsilon(t)\rvert\leq\frac{2}{\pi}\int_0^{\infty}\lvert I(\omega)-I_A(\omega)/\rvert d\omega$
II. Laplace transform	a. $F(s)=\int_0^{\infty}f(t)e^{-st}dt$	$F_A(s)=\frac{f(0)}{s}+\frac{1}{s^2}\cdot\frac{df}{dt}\Big]_{t=0}+\frac{1}{s^2}\sum_{k=1}^{n}\Delta S_{fk}e^{-st_k}$	$F_A(0)=\frac{1}{2}\sum_{k=1}^{n}\Delta S_{fk}t_k^2$	$\lvert\varepsilon(s)\rvert\leq\int_0^{\infty}\lvert f(t)-f_A(t)\rvert dt$ for $Re(s)\geq 0$
	b. $f(t)=\frac{1}{2\pi j}\int_{c-j\infty}^{c+j\infty}F(s)e^{st}ds$	$f_A(t)=\frac{e^{ct}}{2\pi}(-S_{R1}-\sum_{k=1}^{n}\Delta S_{Rk}\cos\omega_k t+\sum_{k=1}^{m}\Delta S_{Ik}\sin\omega_k t)$	$f_A(0)=\frac{1}{2\pi}\sum_{k=1}^{n}\Delta S_{Rk}\omega_k^2$	$\lvert\varepsilon(t)\rvert\leq\frac{e^{ct}}{2\pi}\int_{-\infty}^{+\infty}\lvert R_{FC}-R_{AFC}\rvert d\omega-\frac{e^{ct}}{2\pi}\int_{-\infty}^{+\infty}\lvert I_{FC}-I_{AFC}\rvert d\omega$
III. Mellin transform	$F(s)=\int_0^{\infty}f(t)t^{s-1}dt$	$F_A(s)=\frac{1}{s(s+1)}\sum_{k=1}^{n}\Delta S_{fk}t_k^{s+1}$	$F_A(0)=\sum_{k=1}^{n}\Delta S_{fk}t_k\log t_k$ $F_A(-1)=\sum_{k=1}^{n}\Delta S_{fk}\log t_k$	$\lvert\varepsilon(s)\rvert\leq\int_0^{\infty}\lvert f(t)-f_A(t)\rvert dt$ for $Re(s)\leq 1$
IV. Modified Mellin transform[a]	$F(s)=\frac{1}{t_0^s}\int_{t_0}^{\infty}f(t)t^{s-1}dt$	$F_A(s)=-\frac{f(t_0)}{s}+\frac{t_0}{s(s+1)}\cdot\frac{df}{dt}\Big]_{t=t_0}+\frac{1}{s(s+1)}\sum_{k=1}^{n}\Delta S_{fk}\left(\frac{t_k}{t_0}\right)^s t_k$	$F_A(0)=-f(t_0)+\sum_{k=1}^{n}\Delta S_{fk}t_k\log\frac{t_k}{t_0}$ $F_A(-1)=-f(t_0)+t_0\sum_{k=1}^{n}\Delta S_{fk}\log\frac{t_k}{t_0}$	$\lvert\varepsilon(s)\rvert\leq\frac{1}{t_0}\int_{t_0}^{\infty}\lvert f(t)-f_A(t)\rvert dt$ for $Re(s)\leq 1$
V. Hankel transform of zero order	$F(p)=\int_0^{\infty}f(t)tJ_0(pt)dt$	$F_A(p)=\frac{1}{p^2}\sum_{k=1}^{n}\Delta S_{fk}\left[-t_kJ_0(pt_k)+\frac{2}{p}\sum_{\nu=0}^{\infty}J_{2\nu+1}(pt_k)\right]$	$F_A(0)=\frac{1}{6}\sum_{k=1}^{n}\Delta S_{fk}t_k^3$	$\lvert\varepsilon(p)\rvert\leq\int_0^{\infty}\lvert f(t)-f_A(t)\rvert tdt$

[a] See reference 1.

6. APPLICATION TO OTHER TRANSFORMATIONS

The expressions which have been derived for the Fourier, Laplace, Mellin, and Hankel transforms are presented in Table I along with their corresponding initial value and error bound expressions. Needless to say, the functions, their approximations, and the variables t, p, and s must satisfy the restrictions that hold for each transform. The expressions for the initial values are derived on the assumption that the behavior of the function $f(t)$ as t increases beyond all bounds is such that

$$[f(t)L(t)]_{t=\infty}=0,$$

where $L(t)$ is defined by the indefinite integral

$$L(t)=\int H(0,t)dt.$$

In obtaining the approximate expression for the inverse Fourier transform (Ib) the complex function, $F(j\omega)$, is first broken up into its real and imaginary

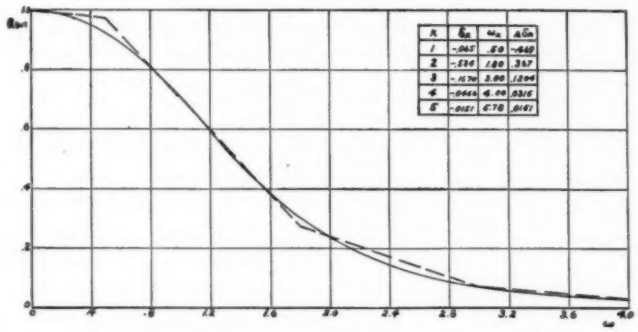


FIG. 2. Approximation of $R(\omega)$ for the first example.

parts:

$$F(j\omega)=R(\omega)+jI(\omega).$$

These parts are then approximated over the interval, $0\leq\omega<\infty$, making use of the even and odd characteristics of $R(\omega)$ and $I(\omega)$, respectively.

In those cases wherein integration in the complex plane is required such that the path of integration is not along the real or imaginary axis, the method is exactly the same. For instance, when the inverse Laplace transform is applied to those s functions whose poles all lie in the left half s plane, the path of integration may be taken along the imaginary axis and the corresponding inverse transform becomes that given on line Ib, Table I. However when the s function has poles in the right half s plane, then the path of integration must be taken to the right of all these poles. That is, this path may be taken from $c-j\infty$ to $c+j\infty$ parallel to the imaginary axis such that c is greater than the real parts of all the poles. In this case the approximate inverse Laplace transform becomes that given by the expression on line IIb, Table I. The straight line approximations

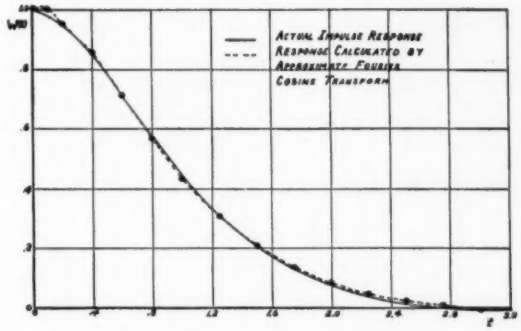


FIG. 3. Comparison of the actual and approximate responses for the first example.

are made to the values that the real and imaginary parts of the s function assume along this path.

All of these expressions are discussed in greater detail in reference 13.

ACKNOWLEDGMENTS

The writer is indebted to Professor James H. Mulligan Jr., Professor Charles F. Rehberg, and Professor Irving F. Ritter of New York University for their many suggestions and criticisms extended throughout the development of this work.

APPENDIX

Examples of the Use of the Approximate Transforms

1. For illustration of the use of expression (9), the real part of a system function given by the following expression:

$$R(\omega)=\frac{1}{0.1223\omega^4+0.300\omega^2+1.000},$$

was approximated by five straight-line segments as shown in Fig. 2. The exact expression for the corresponding impulse response is

$$W(t)=-1.877e^{-1.43t}\cos(0.903t+2.134).$$

The approximate expression obtained by use of expression (9) is

$$W_A(t)=-\frac{2}{\pi t^2}(-0.065-0.469\cos 0.500t+0.367\cos 1.800t+0.1204\cos 3.00t+0.0315\cos 4.00t+0.0151\cos 5.78t),$$

and its initial value is

$$W_A(0)=1.008.$$

These responses are compared in Fig. 3.

[13] A. H. Zemanian, "Investigation of the transient response of linear systems," New York University, doctoral dissertation, 1953, Chapter 5.

• See reference 1.

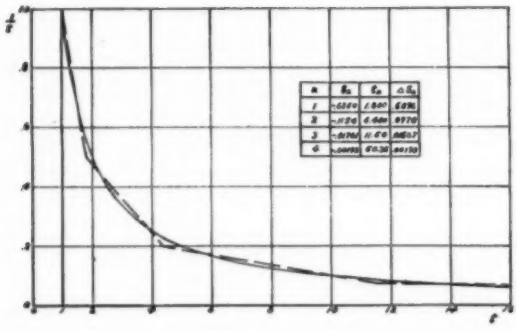
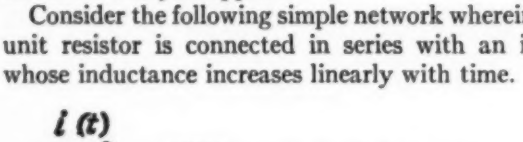


FIG. 4. Approximation of $1/t$ for the second example.

2. As an example of the use of the approximate modified Mellin transform, a problem will be considered wherein the system function of a network is known but the input function is such that its exact transform cannot be determined. To obtain the output, the transform of the input may be determined approximately and then the usual analysis applied.

Consider the following simple network wherein a fixed unit resistor is connected in series with an inductor whose inductance increases linearly with time.



$$v(t)=\frac{1}{t}.$$

The voltage $v(t)$ is impressed at $t=1$ and the resulting current $i(t)$ is to be determined.

This problem yields the differential equation,

$$t di/dt+2i=1/t,$$

and the initial condition,

$$\text{at}\quad t=1,\quad i=0.$$

The exact expression for the response, holding for $t\geq 1$, is

$$i=\frac{1}{t}\left(1-\frac{1}{t}\right).$$

To obtain the approximate solution, $v(t)$ was approximated by four straight line segments as shown in Fig. 4 and the resulting approximate transform for this voltage was found to be

$$V_A(s)=\frac{1}{s}-\frac{0.6250}{s(s+1)}+\frac{1}{s(s+1)}$$
$$\times[0.5096(1.800)^{s+1}+0.0978(4.400)^{s+1}$$
$$+0.01567(11.50)^{s+1}+0.00193(50.35)^{s+1}].$$

Multiplying this expression through by the Mellin transform admittance for this network, $1/(2-s)$, and then taking the inverse transform, the approximate expression for the current is obtained.

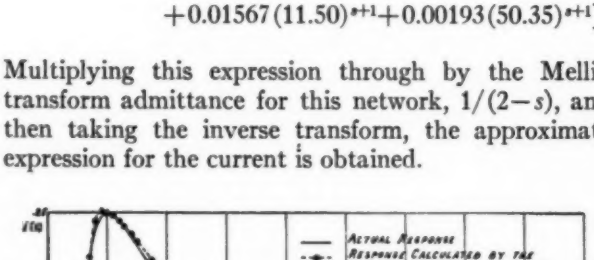


FIG. 5. Comparison of actual and approximate responses for the second example.

$$\text{For } 1.000\leq t\leq 1.800;\ i_A(t)=0.8125-0.2083t-0.6042\frac{1}{t^2};$$

$$\text{for } 1.800\leq t\leq 4.400;\ i_A(t)=0.3538-0.03846t-0.1092\frac{1}{t^2};$$

$$\text{for } 4.400\leq t\leq 11.50;\ i_A(t)=0.1388-0.00587t+1.279\frac{1}{t^2};$$

$$\text{for } 11.50\leq t\leq 50.35;\ i_A(t)=0.04859-0.00064t+5.254\frac{1}{t^2};$$

$$\text{and for } 50.35\leq t;\quad i_A(t)=41.02\frac{1}{t^2}.$$

These results are compared in Fig. 5.

Letters to the Editor

A New Magnetostrictive Material

H. E. STAUSS AND G. SANDOZ
U. S. Naval Research Laboratory, Washington 25, D. C.
(Received September 30, 1953)

MAGNETOSTRICTION, or the change in dimensions of a body placed in a magnetic field, has been observed in ferromagnetic metals and in crystalline ferrites. In addition to these two classes of materials, magnetostriction can be expected to occur in bodies composed of ferromagnetic particles or powders embedded in a matrix that would permit relative motion of the particles. Apparently no search has been made for such materials. In them a change in length can be anticipated either if the particles rotate or if they develop mutual attractions. Either positive or negative magnetostriction could be developed.

The possibility that magnetostriction may occur in a material composed of powdered particles in an appropriate matrix has been examined in this laboratory. Flakes of electrolytic iron were incorporated in a solid, rubbery matrix to the extent of fifty volume percent. The experimental method used for measuring magnetostriction has already been described[1,2] and consisted essentially of putting a block of the specimen in a completed magnetic circuit with an air gap to prevent pressure from the magnet poles on the specimen. Change in dimension was measured by means of a resistance strain gauge.[3] The block was 1 cm×3 cm ×3 cm, and the gap between specimen and pole face was 0.3 cm. In the absence of the specimen the applied magnetic field was 3600 oersteds; no effort was made to measure the apparent induction of the body or the influence of different fields upon the dimensional changes.

Under the conditions described, this particular specimen increased in length by 8×10^{-5} cm/cm. A corresponding decrease was observed when the specimen and attached gauge were turned 90° in the field. At saturation of the individual particles, their individual magnetostrictions would have been about 0.4×10^{-5} cm/cm. If the change in length of the specimen was true magnetostriction, it showed that the flakes probably had rotated in the applied field. To account for the increase in length, the rotation would have had to be some tenths of a minute if it is assumed that the average initial direction was 45° to the applied field.

Consideration must be given in this experiment to the possibility that the extension observed was the result of an attraction across the gap between the magnet pole face and the specimen surface. Whereas use of this experimental method[2] with metals has led to values of magnetostriction comparable to those obtained by other experimenters, the experience with metals does not give conclusive evidence with respect to a resilient material, such as the specimen, which contained a dispersion of magnetic particles. If the specimen can be assumed to have been comparable to a ferromagnetic metal so far as interaction with a field was concerned, the force acting on an individual pole would have been relatively small because of the low apparent induction. The strain, however, could have been large because of the low value of Young's modulus of rubbery materials. If it is assumed that the stress developed across the gap between the specimen face and the magnet pole lay between 0 and $(B-H)B/8\pi$; that the permeability was about 10; and that Young's modulus[4] was about 500 lb/in.,[2] the strain produced would have been between 0 and 1.3 cm/cm, the latter figure being far greater than the measured value of 8×10^{-5} cm. The relationship between the observed extension and possible pole effects is thus far from certain.

The result reported here shows that a material with large positive change in length in a magnetic field was produced by incorporating a ferromagnetic powder in a resilient matrix. Attention is called to material of this type in case it should appear to have practical value.

[1] K. Hoselitz and M. McCaig, Proc. Phys. Soc. (London) **B62**, 163 (1949).
[2] H. E. Stauss and G. Sandoz, J. Metals **4**, 1342 (1952).
[3] J. E. Goldman, Phys. Rev. **72**, 529 (1947).
[4] Walter E. Burton, *Engineering with Rubber* (McGraw-Hill Book Company, Inc., New York, 1949), p. 22.

Change of the Isotopic Abundance Ratio within a Sphere Due to Diffusion*

J. T. BRACKEN AND F. E. SENFTLE
U. S. Geological Survey, Washington 25, D. C.
(Received October 5, 1953)

WHEN diffusion occurs in a solid system containing two or more isotopes of a given element, a change in the isotopic ratio within the body would be expected because of the difference in mass of the isotopes. For some geological problems it is desired to obtain an indication of the average isotopic fractionation within a body, such as a crystal, due to diffusion of isotopes of different masses. To reduce the problem to simplicity an ideal case is assumed.

A hypothetical homogeneous sphere of solvent material is embedded in an infinite homogeneous medium of the same solvent material. Within the sphere at time t equals zero, all of the solute atoms are homogeneously distributed, i.e., the sphere is treated as an instantaneous source. Diffusion of the solute proceeds into the exterior medium at a uniform rate in all directions, and the diffusion coefficient D is assumed to be independent of concentration and the same both inside and outside of the sphere.

The concentration C_r as a function of the radius a of the hypothetical sphere, the time t, and the diffusion coefficient D at any distance $r<a$, is

$$C_r=\frac{3Q_0}{4a^3\pi^{\frac{1}{2}}}\left\{\int_0^{[r+a]/[2(Dt)^{\frac{1}{2}}]}\exp(-y^2)dy-\int_0^{[r-a]/[2(Dt)^{\frac{1}{2}}]}\exp(-y^2)dy-\frac{(Dt)^{\frac{1}{2}}}{r}\left(\exp\left[\frac{-(r-a)^2}{4Dt}\right]-\exp\left[\frac{-(r+a)^2}{4Dt}\right]\right)\right\}, \quad (1)$$

where Q_0 is the original quantity of the diffusing substance (Barrer, 1951; Carslaw and Jaeger, 1947).

If we now consider Q as the total quantity of an isotope of a given mass within the sphere, then

$$Q=4\pi\int_0^a r^2C_r dr. \quad (2)$$

Integrating and substituting (1) in (2), one obtains

$$Q=\frac{3Q_0}{a^3\sqrt{\pi}}\left(\frac{\pi^{\frac{1}{2}}}{2}\left(\int_0^a r^2\operatorname{erf}\frac{r+a}{2(Dt)^{\frac{1}{2}}}dr-\int_0^a r^2\operatorname{erf}\frac{r-a}{2(Dt)^{\frac{1}{2}}}dr\right)+(Dt)^{\frac{1}{2}}\left\{\int_0^a r\exp\left[\frac{-(r+a)^2}{4Dt}\right]-\int_0^a r\exp\left[\frac{-(r-a)^2}{4Dt}\right]\right\}\right). \quad (3)$$

Simplifying, it can be shown that

$$Q=Q_0\left\{\operatorname{erf}\frac{a}{(Dt)^{\frac{1}{2}}}+\frac{2}{\pi^{\frac{1}{2}}}\left(\frac{Dt^{\frac{1}{2}}}{a}\right)^2\left[1-\exp\left(-\frac{a^2}{Dt}\right)\right]-\frac{(Dt)^{\frac{1}{2}}}{a\sqrt{\pi}}\left[3-\exp\left(-\frac{a^2}{Dt}\right)\right]\right\}. \quad (4)$$

This equation satisfies the boundary conditions $Q=Q_0$ for D or $t=0$; and $Q=0$ when D or $t=\infty$; and Q has values of $Q_0\geq Q\geq 0$ for all positive values of Dt and a.

Since for all practical purposes $D_L=(M_H/M_L)^{\frac{1}{2}}D_H=kD_H$, where the subscripts L and H refer to the masses M of light and heavy isotopes, an expression for the isotopic fractionation factor within a spherical body as a function of a, D, and t can be obtained.

Thus

$$f=\frac{Q_L}{Q_H}\cdot\frac{Q_{H_0}}{Q_{L_0}}=\frac{\operatorname{erf}\frac{a}{(kDt)^{\frac{1}{2}}}+\frac{2}{\pi^{\frac{1}{2}}}\left[\frac{(kDt)^{\frac{1}{2}}}{a}\right]\left[1-\exp\left(-\frac{a^2}{kDt}\right)\right]-\frac{(kDt)^{\frac{1}{2}}}{a\sqrt{\pi}}\left[3-\exp\left(-\frac{a^2}{kDt}\right)\right]}{\operatorname{erf}\frac{a}{(Dt)^{\frac{1}{2}}}+\frac{2}{\pi^{\frac{1}{2}}}\left(\frac{(Dt)^{\frac{1}{2}}}{a}\right)\left[1-\exp\left(-\frac{a^2}{Dt}\right)\right]-\frac{(Dt)^{\frac{1}{2}}}{a\sqrt{\pi}}\left[3-\exp\left(-\frac{a^2}{Dt}\right)\right]}. \tag{5}$$

The mean percent isotope separation, $\bar{S}$, within the sphere, is defined as

$$\bar{S}=(1-f)\cdot 10^2. \tag{6}$$

The authors wish to thank Roland G. Henderson of the U. S. Geological Survey for aid in developing the equations. This work has been sponsored by the Division of Research of the U. S. Atomic Energy Commission.

* Publication authorized by the Director, U. S. Geological Survey, Washington 25, D. C.

Radioactive-Shadowing Technique and Note on the Deposition of Metal Crystals from Atomic Beams

M. Simnad, Ling Yang, and G. M. Pound
Metals Research Laboratory, Carnegie Institute of Technology, Pittsburgh, Pennsylvania
(Received September 12, 1953)

EXPERIMENTS have been carried out to find the effect of oblique metal vapor beams upon the distribution of the deposits on the targets. It was believed that such a method would accentuate the effect of surface-roughness factors upon the uniformity of the deposit formed from the vapor.

A high vacuum vapor effusion apparatus, which will be described elsewhere,[1] was designed to produce a collimated, radioactive silver vapor beam. The target could be placed in the path of the vapor beam at any predetermined angle. Two targets were made of copper disks, 1 cm diameter ×1 mm thick and 2 cm diameter ×1 mm thick. A series of concentric steps 1/1000 in. high were machined on their surfaces, the central area being the highest part of the surface. The first target was placed at an angle of about 10° to the vapor beam. The second target was placed at an angle of about 5° to the beam and a copper wire, ½-mm diameter, was placed in the path of the beam as shown in Fig. 1. A quantity of radioactive silver, Ag^{110}, corresponding to an average thickness of 500A was deposited upon the targets. The latter were then placed on no-screen x-ray film, and their autoradiographs were taken. The exposure given was four days.

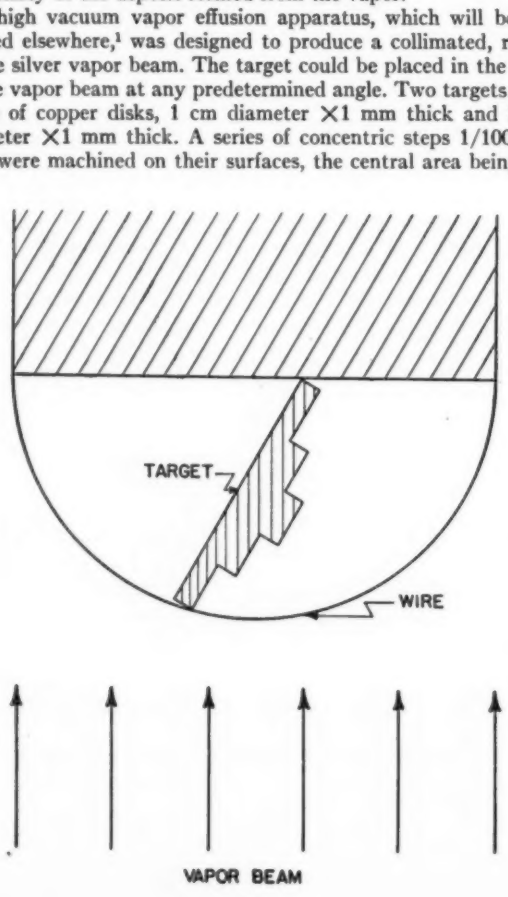


Fig. 1.

The autoradiographs of these targets are shown in Figs. 2 and 3, which show up the surface irregularities very distinctly. The dark arcs correspond to the step-face regions facing the vapor beam. The white areas correspond to the parts that were shielded from the beam by the steps. Also, the roughly machined surfaces show evidence of nonuniform deposition even on the flat areas between the steps.

The quantitative relationships between collision frequency of atoms from the beam with unit substrate surface, concentration of adsorbed atoms, and nucleation rate are somewhat different from those given for heterogeneous nucleation from ordinary vapor[2] and will be discussed in a forthcoming publication.[1] However, for the present, it is sufficient to say that deposition rate is strongly dependent on the frequency per unit area with which atoms from the uniform beam strike the substrate surface. This frequency is directly proportional to the sine of the angle between the portion of surface in question and the beam. Accordingly, surface irregularities will give rise to an uneven deposition. The regions of the surface that happen to lie most nearly perpendicular to the beam will have the greatest number of atoms striking them, and deposition will be most rapid on these areas.

In the autoradiograph of the part of the surface under the wire there is evidence of a deposit, the amount of which increases with the distance of the wire from the surface. This condition would appear to be due to diffraction of the beam around the wire. However, the de Broglie wavelength corresponding to the mean translational kinetic energy of silver atoms in the beam is only

$$\lambda=h/(mkT)^{\frac{1}{2}}=10^{-9}\text{ cm}.$$

The Fresnel diffraction around the wire at the surface, L cm distant from the wire in the direction of the beam, is given approximately by

$$\delta=(\lambda L/2)^{\frac{1}{2}}.$$

At the maximum L of 2 cm, δ is only about 10^{-4} cm. The deflection around the wire shown in Fig. 3 is of the order of 1000-fold greater than this. Therefore, diffraction does not seem to be a tenable explanation. Also, from the geometry of the apparatus that produces the atomic beam, it is possible that imperfect collimation

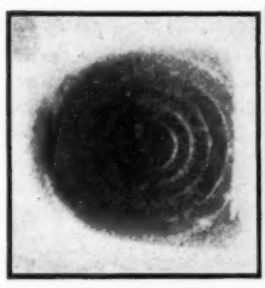


Fig. 2.

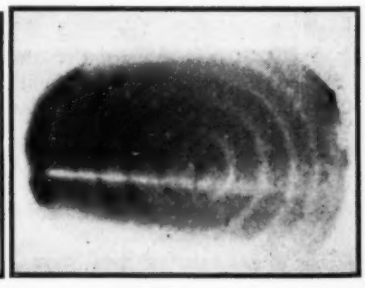


Fig. 3.

could produce the observed result. Further, surface diffusion cannot account for the effects seen in Fig. 3. Experiments are now being carried out to determine the nature of this apparent diffraction.

This radioactive shadowing technique shows promise of being a useful method for studying small surface inhomogeneities. Quantitative studies of heterogeneous condensation of metal crystals from radioactive atomic beams are being carried out and will be presented later. Reference should be made to work in this field by Cockcroft[3] and Lennard,[4] which will be considered in detail in future communications.

Support of this work has been given by a research program sponsored by the U. S. Office of Naval Research.

[1] Yang, Birchenall, Pound, and Simnad, *Heterogeneous Condensation of Metal Crystals from the Vapor Phase* (to be published).
[2] J. H. Hollomon and D. Turnbull, *Progress in Metal Physics, IV* (Interscience Publishers, Inc., New York, 1953), pp. 333–388.
[3] J. D. Cockcroft, Proc. Roy. Soc. (London) **A119**, 293 (1928).
[4] J. J. E. Lennard, Proc. Phys. Soc. (London) **49**, 140 (1937).

Comments on "Diffraction of Electromagnetic Waves by an Aperture in a Large Screen"

J. H. CRYSDALE
Radio Physics Laboratory, Defence Research Board, Ottawa, Canada
(Received October 26, 1953)

A RECENT paper by G. Bekefi[1] is concerned with the problem of the diffraction fields of an irradiated aperture in a perfectly-conducting screen. The significance of some of Bekefi's assumptions and approximations is discussed in this letter. In addition, it will be shown that Bekefi's approximate formula can be derived from a simple and straightforward argument based on the physical aspects of the problem.

In order to facilitate the analysis of Bekefi's paper, it seems advisable to review the application of the electric Hertz vector[2] to the problem of an irradiated aperture in a plane, perfectly-conducting screen of vanishing thickness. As in Bekefi's paper, it is assumed that the screen lies in the plane $z=0$ and that the primary sources are located in the region $z<0$; consequently, the field is due to primary and induced sources in the region $z\leq 0$. Harmonic time variation $e^{j\omega t}$ is implicitly assumed throughout. The symbol $\mathbf{\Pi}_i$ will be used to denote the electric Hertz vector resulting from the sources in the region $z<0$; the electric Hertz vector resulting from the sources induced on the surfaces of the screen will be represented by $\mathbf{\Pi}_s$.

The electric Hertz vector is defined by the differential equation

$$\nabla^2\mathbf{\Pi}+k^2\mathbf{\Pi}=\frac{j\mathbf{J}}{\epsilon\omega}, \tag{1}$$

where $\mathbf{J}$ is the volume electric current density. Consequently,

$$\mathbf{\Pi}_s=\frac{-j}{4\pi\epsilon\omega}\int_s \mathbf{K}\frac{e^{jkr}}{r}da, \tag{2}$$

where $\mathbf{K}$ is the surface electric current density and the integration is over both surfaces of the plane screen. Since all components of the electric Hertz vector are continuous through the surface of a perfect conductor,

$$\mathbf{\Pi}=0, \tag{3}$$

on both surfaces of the screen. Also, because of the assumption of vanishing screen thickness, it is evident from Eq. (2) that $\mathbf{\Pi}_s$ is symmetrical with respect to the plane $z=0$; accordingly,

$$\frac{\partial\mathbf{\Pi}_s}{\partial z}=0 \tag{4}$$

in the aperture. Consequently,

$$\frac{\partial\mathbf{\Pi}}{\partial z}=\frac{\partial\mathbf{\Pi}_i}{\partial z} \tag{5}$$

in the aperture. It is evident that $\mathbf{\Pi}_s$ can be written in the form

$$\mathbf{\Pi}_s=\Pi_{sx}\mathbf{a}_x+\Pi_{sy}\mathbf{a}_y, \tag{6}$$

where $\mathbf{a}_x$ and $\mathbf{a}_y$ are unit vectors parallel to the x and y axes. It will be assumed that the distribution of sources in the region $z<0$ is such that $\mathbf{\Pi}_i$ can be expressed in the form

$$\mathbf{\Pi}_i=\Pi_{ix}\mathbf{a}_x. \tag{7}$$

Consequently, the total Hertz vector $\mathbf{\Pi}$ is

$$\begin{aligned}\mathbf{\Pi}&=(\Pi_{ix}+\Pi_{sx})\mathbf{a}_x+\Pi_{sy}\mathbf{a}_y,\\ &=\Pi_x\mathbf{a}_x+\Pi_y\mathbf{a}_y.\end{aligned} \tag{8}$$

It is not difficult to deduce from the properties of the fields of elementary electric charges and currents that the sources induced in the surfaces of the screen do not contribute to either the normal component of the electric field in the aperture or the tangential components of the magnetic field in the aperture.

The electric and magnetic intensities can be determined from the electric Hertz vector with the relations

$$\mathbf{E}=\nabla\nabla\cdot\mathbf{\Pi}+k^2\mathbf{\Pi}, \tag{9a}$$

$$\mathbf{H}=\frac{-k^2}{j\omega\mu}\nabla\times\mathbf{\Pi}. \tag{9b}$$

From Eq. (8) and (9) it follows that

$$E_x=\frac{\partial^2\Pi_x}{\partial x^2}+\frac{\partial^2\Pi_y}{\partial x\partial y}+k^2\Pi_x, \tag{10a}$$

$$E_y=\frac{\partial^2\Pi_x}{\partial x\partial y}+\frac{\partial^2\Pi_y}{\partial y^2}+k^2\Pi_y, \tag{10b}$$

$$E_z=\frac{\partial^2\Pi_x}{\partial x\partial z}+\frac{\partial^2\Pi_y}{\partial y\partial z}, \tag{10c}$$

$$H_x=\frac{k^2}{j\omega\mu}\frac{\partial\Pi_y}{\partial z}, \tag{10d}$$

$$H_y=\frac{-k^2}{j\omega\mu}\frac{\partial\Pi_x}{\partial z}, \tag{10e}$$

$$H_z=\frac{k^2}{j\omega\mu}\left(\frac{\partial\Pi_x}{\partial y}-\frac{\partial\Pi_y}{\partial x}\right). \tag{10f}$$

It is well known that the electric current on the surfaces of the screen must have both x and y components in order to satisfy the boundary conditions on the rim of the aperture;[3-6] therefore, a theory in terms of Π_x alone could not be exact. Bekefi pointed out that the use of only the x component of the electric Hertz vector imposes the not necessarily true restriction that $H_x=0$ at every point in space. It is evident from Eqs. (10a–f) that such a solution can result in erroneous predictions of E_x, E_y, E_z and H_z as well.

Bekefi also discussed the solution of the problem in terms of the magnetic Hertz vector.[2] It is known that the modifications in the field quantities due to the presence of an aperture in an infinite, plane, perfectly-conducting, vanishingly-thin screen can be expressed in terms of fictitious distributions of magnetic currents and charges in the aperture.[6] Let $\mathbf{\Pi}_s^*$ denote the magnetic Hertz vector due to these magnetic sources. It is evident that $\mathbf{\Pi}_s^*$ is symmetrical with respect to the plane $z=0$; therefore,

$$\frac{\partial\mathbf{\Pi}_s^*}{\partial z}=0 \tag{11}$$

at all points in the plane $z=0$ not in the aperture. If it is assumed that the source distribution in the region $z<0$ is the same in the presence and absence of the aperture in the screen, the electric and magnetic intensities are

$$\mathbf{E}=\mathbf{E}_0-\frac{k^2}{j\omega\epsilon}\nabla\times\mathbf{\Pi}_s^*, \quad (z\leq 0), \tag{12a}$$

$$=\frac{k^2}{j\omega\epsilon}\nabla\times\mathbf{\Pi}_s^*, \quad (z\geq 0), \tag{12b}$$

$$\mathbf{H}=\mathbf{H}_0-(\nabla\nabla\cdot\mathbf{\Pi}_z^*+k^2\mathbf{\Pi}_z^*), \quad (z\leq 0), \tag{12c}$$

$$=\nabla\nabla\cdot\mathbf{\Pi}_z^*+k^2\mathbf{\Pi}_z^*, \quad (z\geq 0), \tag{12d}$$

where $\mathbf{E}_0$ and $\mathbf{H}_0$ are the electric and magnetic intensities in the region $z\leq 0$ when the aperture is not present in the screen. Expressions for the rectangular components of the field quantities analogous to Eqs. (10a–f) can be derived from Eqs. (12a–d) when it is noted that

$$\mathbf{\Pi}_z^*=\Pi_{zz}^*\mathbf{a}_z+\Pi_{zy}^*\mathbf{a}_y. \tag{13}$$

In particular, the expressions for E_z, H_z, and H_y in the region $z\geq 0$ are

$$E_z=\frac{-k^2}{j\omega\epsilon}\left(\frac{\partial\Pi_{zz}^*}{\partial y}-\frac{\partial\Pi_{zy}^*}{\partial x}\right), \tag{14a}$$

$$H_z=\frac{\partial^2\Pi_{zz}^*}{\partial x^2}+\frac{\partial^2\Pi_{zy}^*}{\partial x\partial y}+k^2\Pi_{zz}^*, \tag{14b}$$

$$H_y=\frac{\partial^2\Pi_{zz}^*}{\partial x\partial y}+\frac{\partial^2\Pi_{zy}^*}{\partial y^2}+k^2\Pi_{zy}^*. \tag{14c}$$

Since the electric charges and currents on the surfaces of the screen do not contribute to E_z, H_x, and H_y in the aperture, it is evident that the application of Eqs. (14a–c) to points in the aperture yields a set of conditions on $\mathbf{\Pi}_z^*$ in the aperture in terms of the sources in the region $z<0$.

Bekefi states that it is possible to construct an electromagnetic field from the vector component Π_y^* which fulfills the boundary requirements

$$\Pi_y^*=\Pi_{iy}^* \tag{15a}$$

in the aperture,

$$\frac{\partial\Pi_y^*}{\partial z}=0 \tag{15b}$$

on the screen. He points out that the field calculated from these boundary requirements is not in agreement with experimental measurements. It would seem that the correct explanation for this disagreement is that there is no fundamental justification for the boundary condition of Eq. (15a) even though it is consistent with Eqs. (14a–c). The condition of Eq. (15b) is justified by Eq. (11). On the other hand, in Bekefi's analysis in terms of the electric Hertz vector, his assumptions regarding the value of Π_z on the surface of the screen and the value of $\partial\Pi_z/\partial z$ in the aperture are justified by Eqs. (3) and (5) of this letter.

In Sec. III of his paper, Bekefi states that Π_z for a uniformly illuminated circular aperture can be expressed in the form

$$\Pi_z(\rho',z')=\int_0^\infty J_0(\lambda\rho')f(\lambda a)e^{-\mu z'}\frac{\lambda d\lambda}{\mu}. \tag{16}$$

ρ' and z' are coordinates in a cylindrical coordinate system. The definitions of the remaining symbols can be found in Bekefi's paper. Actually, the integral representation of Eq. (16) is not applicable to this problem since its use implies that Π_z is independent of ϕ', where ϕ' is the angular coordinate in the cylindrical coordinate system (ρ',ϕ',z'). It is evident from Eq. (2) that, in general, $\Pi_z(\rho',\phi',z')$ would not be independent of ϕ' since, in general,[4,5]

$$\frac{\partial K_z(\rho,\phi,0)}{\partial\phi}\neq 0. \tag{17}$$

It will now be shown that Bekefi's approximate formula for Π_z can be derived with a simple physical argument. The assumptions are:

1. The field incident on the screen is uniform, plane and linearly polarized, and can be described by the equations

$$\mathbf{E}=E_c e^{-jkz}\mathbf{a}_x, \tag{18a}$$

$$\mathbf{H}=\frac{E_c}{Z}e^{-jkz}\mathbf{a}_y, \tag{18b}$$

where Z is the free space wave impedance. E_c is independent of the coordinates.

2. The currents induced on both surfaces of the screen are the currents predicted by the principles of geometrical optics.

From assumption 1 and Eq. (9a), it follows that

$$\mathbf{\Pi}_i(x',y',z')=\frac{E_c}{k^2}\exp(-jkz')\mathbf{a}_x. \tag{19}$$

From assumption 2, it can be found that

$$\mathbf{K}=\frac{2E_c}{Z}\mathbf{a}_x \tag{20a}$$

on the illuminated side of the screen,

$$\mathbf{K}=0 \tag{20b}$$

on the dark side of the screen. Therefore,

$$\mathbf{\Pi}_z(x',y',z')=\frac{-j2E_c\mathbf{a}_x}{4\pi\omega Z}\int_{s-}\frac{e^{-jkr}}{r}da, \tag{21}$$

where s_- denotes the illuminated side of the screen. The integral in Eq. (21) can be evaluated following the method of Schoch;[7] as is usual with such integrals, it is convenient to assume that k is slightly complex. If $(\rho_1,\phi_1,0)$ denotes a point which lies in the aperture, it follows that

$$\mathbf{\Pi}_z(\rho_1,\phi_1,z_1)=\frac{-E_c\mathbf{a}_x}{2\pi k^2}\int_0^{2\pi}e^{-jkr_0}d\alpha. \tag{22}$$

The symbols r_0 and α are defined in Bekefi's paper. Consequently,

$$\mathbf{\Pi}(\rho_1,\phi_1,z_1)=\frac{E_c\mathbf{a}_x}{k^2}\left[e^{-jkz_1}-\frac{1}{2\pi}\int_0^{2\pi}e^{-jkr_0}d\alpha\right]. \tag{23}$$

The reduction of Eq. (23) to Bekefi's Eq. (22) is obvious. Schoch's method can also be used to determine the corresponding expression for $\mathbf{\Pi}(\rho_2,\phi_2,z_2)$ where the point $(\rho_2,\phi_2,0)$ lies in the screen.

In conclusion, it might be pointed out that Bouwkamp[8,9] has criticized the papers by Bethe[6] and Copson[10] referred to by Bekefi in his introduction. Copson at first disagreed with Bouwkamp's criticisms,[11] but later acknowledged that they were correct.[12]

[1] G. Bekefi, J. Appl. Phys. **24**, 1123 (1953).
[2] J. A. Stratton, *Electromagnetic Theory* (McGraw-Hill Book Company, Inc., New York, 1941), pp. 28, 430.
[3] C. J. Bouwkamp, Physica **12**, 467 (1946).
[4] J. Meixner, Ann. Physik **6**, 1 (1949).
[5] J. Meixner and W. Andrejewski, Ann. Physik **7**, 157 (1950).
[6] H. A. Bethe, Phys. Rev. **66**, 163 (1944).
[7] A. Schoch, Akust. Z. **6**, 318 (1941).
[8] C. J. Bouwkamp, Philips Research Reports **5**, 321 (1950).
[9] C. J. Bouwkamp, Math. Revs. **8**, 179 (1947).
[10] E. T. Copson, Proc. Roy. Soc. (London) **A186**, 100 (1946).
[11] E. T. Copson, Proc. Roy. Soc. (London) **A202**, 277 (1950).
[12] E. T. Copson, Math. Revs. **12**, 774 (1951).

Momentum Transfer in Sputtering by Ion Bombardment

GOTTFRIED K. WEHNER
Electronic Components Laboratory, Wright Air Development Center, Wright-Patterson Air Force Base, Ohio

(Received November 12, 1953)

THE most widely accepted sputtering theory is the evaporation theory, originated by Hittorf[1] and developed particularly by v. Hippel[2] and Townes.[3] For a sputtered wire, Fetz[4] found that the rate of sputtering increases as the diameter of the wire decreases. He ascribes this fact to an increase of the sputtering rate when ions strike the target under oblique angles. This result would not be explicable in terms of the evaporation theory but supports the momentum transfer theoreies as developed by Stark[5] and Holst,[6] Lamar and Compton,[7] and Kingdon and Langmuir.[8] Massey and Burhop state,[9] "It is evident, however, that more direct experiments are needed on the effect of angle of incidence of the ions in order to establish the validity of Fetz' interpretation."

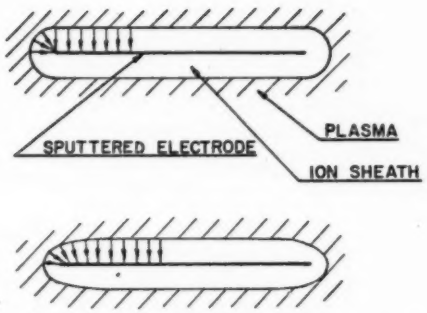


FIG. 1. Cross section of metal strip under ion bombardment.

Some simple and conclusive experiments, confirming the results of Fetz and revealing other strong evidence in favor of the momentum-transfer theories, can be made in a plasma of high density (10^{12} per cm^3) and low gas pressure (1 micron or less). Such a plasma provides ions with well defined ion velocities and angles of incidence (small ion sheath thickness—no collisions within the ion sheath), a high sputtering rate even at low ion energies, and an undistrubed transfer of the sputtered material to the places of deposition.

When a metal strip is immersed in such a plasma and bombarded with positive ions, most of the material is sputtered not from the edges but from a seam along the edges of the sample as indicated in the cross-section picture of Fig. 2. When sputtering is continued for a long enough time, a strip of material is finally separated from the sheet along the edges. Figure 1 shows a cross section of the original metal strip and its surrounding ion sheath. If the ion sheath had the form of the upper picture, then the edge proper would be attacked most heavily. The actual form of the ion sheath is more elliptical, as in the lower picture, as can be shown by calculating and comparing the sheath thickness for the plane and cylindrical case. The ion paths indicated show why the place of heaviest attack is shifted towards the inside.

The question now arises as to whether this preferential attack of the surface is due alone to the somewhat higher current density in this region or whether in addition to this, sputtering under oblique angles is more efficient. The answer is found by study of the sputtered deposits when the metal strip is placed in the axis of a large glass cylinder which catches the sputtered material. The result is indicated as photographed in Fig. 2. For the purpose of checking where the four separated deposits came from, the experiment was repeated with the metal strip covered on the top side along the right edge with aquadag, which is known to withstand sputtering very well. From the lower photograph of Fig. 2, it can be seen that the opposite (upper left) deposit disappears, clearly indicating that momentum has been transferred from the obliquely incident ions to the sputtered atoms.

Increasing the ion energy and keeping the ion sheath thickness constant by adjusting the plasma density accordingly ($K=KU^{\frac{1}{2}}$) brings the two separate deposits closer and closer together. Finally, at a certain ion energy, which is different and characteristic for every gas-metal combination, these deposits merge together giving just one maximum on each side of the metal strip. The explanation of this fact is that, at low ion energies, sputtering from oblique angle ions is much more efficient and the big center area of the metal strip, which is under perpendicular bombardment, contributes little to the deposits. At higher ion energies, however, the difference between perpendicular and oblique bombarding ions becomes less and less pronounced so that finally the sputtering from the big central area contributes most to the deposits. This fact explains earlier and different results by Seeliger and Sommermeyer,[10] who made their experiments with ion energies between 5000 and 10 000 volts. This, however, indicates that the angle of incidence largely determines the threshold energy (the least amount of energy necessary for separating metal atoms from the surface) and explains why experiments in the normal glow discharge, or with wires which are thin compared with the surrounding ion sheath and therefore undetermined angles of incidence, do not give clearly defined or reliable threshold energies.

The experiments reported confirm not only the results of Fetz, but show in addition that under low ion energy bombardment momentum is transferred to the sputtered atoms. The angle of incidence is an important and previously neglected parameter in the determination of threshold energies. In every metal-gas combination, a different and characteristic minimum threshold energy is necessary to start sputtering at perpendicular incidence. These results cannot be explained by the evaporation theory. The measurement of threshold energies for different metals in Hg and their connection with the heat of formation and the velocity of sound in the metal will be reported in the near future.

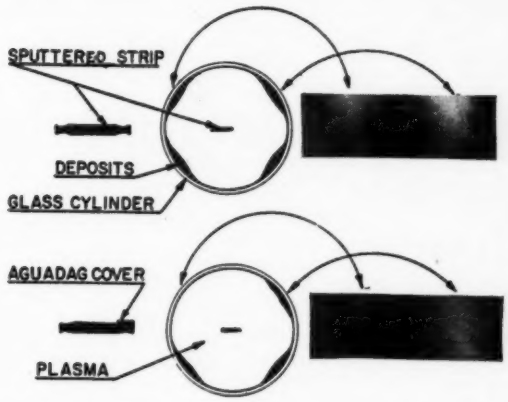


FIG. 2. Deposits on a glass cylinder surrounding the sputtered metal strip. Contact photographs show the deposits on the upper half of the glass cylinder.

[1] W. Hittorf, Ann. Physik **21**, 126 (1884).
[2] v. Hippel, Ann. Physik **81**, 1043 (1926).
[3] C. H. Townes, Phys. Rev. **65**, 319 (1944).
[4] H. Fetz, Z. Physik **119**, 590 (1942).
[5] J. Stark, Z. Elektrochem. **14**, 752 (1908); **15**, 509 (1909).
[6] G. Holst, Physica **4**, 68 (1924).
[7] Lamar and Compton, Science **80**, 541 (1934).
[8] Kingdon and Langmuir, Phys. Rev. **22**, 148 (1923).
[9] Massey and Burhop, *Electronic and Ionic Impact Phenomena* (Oxford University Press, New York, 1952).
[10] Seeliger and Sommermeyer, Z. Physik **93**, 692 (1935).

Rotation of the Microwave Polarization in Liquid Crystals

GLENN A. MANN AND R. D. SPENCE
Department of Physics, Michigan State College, East Lansing, Michigan
(Received November 12, 1953)

THE experiment reported in this note is an outgrowth of some previous observations on the behavior of the microwave dielectric constant of a liquid crystal placed in the static magnetic field. We have shown* that for a compound such as *para*-azoxyanisole in the liquid-crystal phase there is a considerable difference in the complex dielectric constants measured with the microwave electric field parallel and perpendicular to the transverse static magnetic field. For example, for *para*-azoxyanisole at a temperature of 128°C and a frequency 15 000 Mc, the dielectric constants parallel and perpendicular to the static magnetic field are $\epsilon_{\parallel}=(3.79-i0.47)\epsilon_0$, $\epsilon_{\perp}=(3.46-i0.85)\epsilon_0$.

The large difference in the imaginary part of the dielectric constant leads one to expect that if microwaves were allowed to enter the liquid crystal polarized at an angle ψ_0 with respect to the magnetic field, the component polarized perpendicular to the magnetic field would rapidly be attenuated compared with the

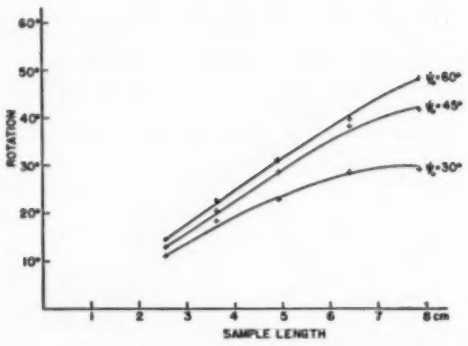


FIG. 1. Rotation of the plane of polarization in *p*-azoxyanisole as a function of the length of sample; ψ_0 is the angle between the static magnetic field and the electric vector of the microwave field. The data for this figure were taken at a temperature of 129°C in a field of 1300 gauss and at a frequency 9 400 Mc.

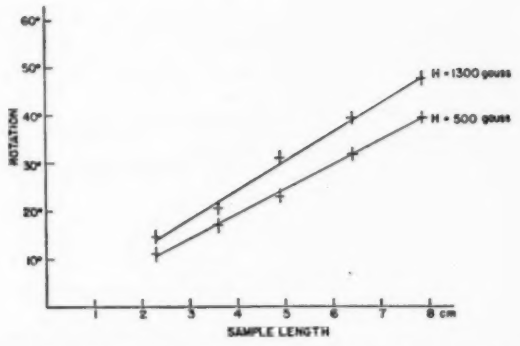


FIG. 2. Rotation as a function of sample length for $\psi_0=60°$ and two different values of the magnetic field. Temperature and frequency are the same as in Fig. 1.

parallel component. To observe this effect, we used a circular wave guide operating in the TE_{11} mode. We chose to use *para*-azoxyanisole since its liquid-crystal phase extends over a fairly wide temperature range (17°C), occurs at low temperatures, and is easy to prepare. The experimentally measured rotations as a function of the length of sample and for different angles of incidence ψ_0 are shown in Fig. 1. These curves are for a field of 1300 gauss and a temperature of 129°C. The ultimate rotations for a given angle of incidence ψ_0 approach ψ_0. This fact simply means that the field eventually becomes polarized along the direction of the static magnetic field if the sample is long enough. Figure 2 shows the rotation for an angle of incidence ψ_0 equal to 60° and for the two fields of 1300 and 500 gauss. At fields above 1300 gauss, the effect is fairly well saturated, and any further increase of the field intensity does not produce any considerable increase in the rotation. Below 400 gauss, the rotation falls off rapidly and becomes somewhat erratic because of the alignment of the molecules by thermal gradients in the column of liquid crystal.

It appears that if this effect is ever to be used in any practical device, two modifications will have to be sought. One needs a compound whose liquid-crystal phase occurs at room temperature and in which the attenuation of the electric field component parallel to the static magnetic field is very small. At present it appears likely that a compound in which the liquid-crystal phase occurs at room temperature can be found. However, it is not clear whether it is possible to find a compound in which the attenuation of the parallel component of the electric field is really small.

* E. F. Carr and R. D. Spence, J. Chem. Phys. (to be published).

Cover Photograph

Climaxing a seven-year series of contracts with the U. S. Signal Corps, The Brush Laboratories Company, Unit of Clevite Corporation of Cleveland, recently harvested this clear and shapely crystal from one of its quartz-growing autoclaves. It is the world's largest and most nearly perfect man-made quartz crystal. The crystal weighs 1014 grams and was grown on a Brazilian quartz seed-plate 3×5 inches by $\frac{1}{8}$ inch thick and weighing 4.5 percent of the present total. The process of growth was carried out over a period of 78 days, at a temperature of 660°F, a pressure of 5000 pounds per square inch, and in the presence of 18 percent sodium carbonate in water. The feed material was small broken chunks from Brazilian quartz crystals.

The crystal is considered very high quality by the radio crystal industry. It contains one small speck (about 10 mm²) of optical twinning, and, on the basis of destructive examination of other synthetic quartz, it is believed to be free from electrical twinning. In short, this crystal is of higher quality, weight for weight, than 99 percent of all natural quartz.

The pilot plant of The Brush Laboratories has lately (August 25) produced another crystal weighing 1395 grams in a run lasting 131 days; its two major surfaces, however, contain a number of grooves extending into the crystal for about $\frac{1}{8}$ of the crystal thickness.

These radiograde crystals were grown in autoclaves having capacities of 50 pounds of synthetic quartz per run of 75 days. Usually, instead of such large crystals, a large number of crystals weighing about 90 grams each are produced, these being of a size much more useful for processing into oscillator plates.

Careful tests by the Signal Corps and by several manufacturers of oscillator units have shown that this synthetic product is satisfactory for use in all frequency control applications now filled by natural quartz. Chemically, it is identical with natural quartz except a little purer.

The work of Giorgio Spezia, Italian mineralogist, in 1905–1908 and Richard Nacken, contemporary German mineralogist, contributed to the success of this project.